Niveau(x) d'enseignement: 0 Numéro EDUQ : 9470109
Centre d'intérêt: 540

L'Evaluation des apprentissages : réflexions, nouvelles tendances et formation / sous la direction de René Hivon. -- [Sherbrooke] : Editions du CRP, 1993.
250 p. ; 23 cm.

"Faculté d'éducation, Université de Sherbrooke". Bibliogr. à la fin des chap.

ISBN 2-920859-69-2

Vedette(s)-secondaire(s): Hivon, René, éd * Université de Sherbrooke. Faculté d'éducation

Descripteur(s): Evaluation de l'apprentissage * Transfert d'apprentissage * Evaluation formative * Psychologie cognitive * Métacognition * Evaluation par l'étudiant * Mesure * Utilisation de l'ordinateur * Test de langage * Apprentissage de l'adulte * Opinion de l'enseignant * Rôle de l'enseignant * Français (Langue) * Evaluation de tests * Validité

Identificateur(s) de sujet: Constructivisme

Contribution des chercheurs qui ont participé à un séminaire dans le cadre du symposium du Réseau international de recherche en éducation et formation qui s'est tenu à Sherbrooke, en octobre 1992, sous le thème spécifique "L'évaluation des apprentissages". Dans une première partie, ils abordent cette question à la lumière des recherches récentes en psychologie cognitive et en évaluation formative. Par la suite, ils traitent de l'utilisation de l'informatique en évaluation formative. Suivent quatre textes qui touchent l'éducateur et l'évaluation: croyances, formation, pratiques. Un dernier texte porte sur le concept de validité dans le contexte des examens du MEQ.

Type de document:

Collectif sous la direction
de RENÉ HIVON

L'ÉVALUATION DES APPRENTISSAGES

Réflexions, nouvelles tendances et formation

Éditions du CRP
Faculté d'éducation
UNIVERSITÉ DE SHERBROOKE

94-70109

Collectif sous la direction de RENÉ HIVON

L'ÉVALUATION DES APPRENTISSAGES : RÉFLEXIONS, NOUVELLES TENDANCES ET FORMATION

Distribution en Europe : Les Éditions ESKA

ISBN 2-920859-69-2

© Éditions du CRP
Faculté d'éducation
Université de Sherbrooke
Québec, Canada

Dépôt légal : 2e trimestre 1993

Bibliothèque nationale du Québec, Montréal
Bibliothèque nationale du Canada, Ottawa

Tous droits de reproduction, même partielle, d'adaptation
ou de traduction réservés.

TABLE DES MATIÈRES

PRÉSENTATION
René Hivon (Université de Sherbrooke) 11

PREMIER CHAPITRE – OBJECTIFS TERMINAUX D'INTÉGRATION
TRANSFERT DES CONNAISSANCES
Jean-Marie De Ketele (Université Louvain-La-Neuve) 15

Introduction 15
1. Inventaire descriptif des pratiques évaluatives 15
1.1 Les pratiques centrées sur les contenus 15
1.2 Les pratiques centrées sur les activités 16
1.3 Les pratiques centrées sur les objectifs spécifiques 16
1.4 Les pratiques centrées sur les réseaux conceptuels 16
1.5 Les pratiques centrées sur une activité ou thème intégrateur 17
1.6 Les pratiques centrées sur les objectifs d'intégration 18
2. Les déterminants de la performance académique 19
3. Évaluation des pratiques évaluatives 20
3.1 Les pratiques centrées sur les contenus 20
3.2 Les pratiques centrées sur les activités 21
3.3 Les pratiques centrées sur les objectifs spécifiques 22
3.4 Les pratiques centrées sur les réseaux conceptuels 23
3.5 Les pratiques centrées sur une activité ou thème intégrateur 23
3.6 Les pratiques centrées sur les objectifs d'intégration 24
Conclusion 25

DEUXIÈME CHAPITRE – L'ÉVALUATION DANS LE PARADIGME CONSTRUCTIVISTE
Jacques Tardif (Université de Sherbrooke) 27

Introduction 27
1. Une définition de l'évaluation 29
2. Quelques principes d'apprentissage 32
3. Quelques caractéristiques de l'évaluation authentique 43
4. Au sujet de la fidélité et de la validité 49
Conclusion 53

TROISIÈME CHAPITRE – ÉVALUATION FORMATIVE DES PROCESSUS D'APPRENTISSAGE : LE RÔLE DES RÉGULATIONS MÉTACOGNITIVES
Linda Allal (Université de Genève) 57

Introduction 57
1. Directions de développement des travaux européens en évaluation formative 57
1.1 Les techniques d'instrumentation 58
1.2 Les fondements théoriques 58
1.3 Les pratiques effectives et leur mise en contexte 59
1.4 L'implication de l'élève 59
2. Les régulations intervenant dans le système didactique 60
2.1 Les régulations orchestrées par l'enseignant 61
2.2 Les régulations gérées par les élèves 64
3. Régulations métacognitives et modalités d'évaluation formative favorisant leur mise en œuvre 66
3.1 Les opérations de régulation métacognitive 66
3.2 Mise en œuvre des régulations et des modalités d'évaluation formative 67
Conclusion 69

QUATRIÈME CHAPITRE – RÉGULATION DES APPRENTISSAGES, CONTEXTE DE MESURE ET ÉVALUATION INFORMATISÉE
Clément Dassa (Université de Montréal) 75

Introduction 75
1. Le contexte actuel de l'évaluation 76
2. La place de l'évaluation formative dans une pédagogie efficace 77
3. L'élargissement de la régulation 79
4. Vers un cadre conceptuel des pratiques de l'évaluation 80
5. Les pratiques évaluatives : entre le projet et les contraintes 83
6. Le besoin d'outillage et les technologies informatiques 86
7. L'importance d'une conception pédagogique des APO 88
8. Diagnostic pédagogique et informatisation 88
9. Le diagnostic informatisé facteur d'intégration 91
Conclusion 92

CINQUIÈME CHAPITRE – UN TEST ADAPTATIF EN LANGUE SECONDE : LA PERCEPTION DES APPRENANTS
Michel Laurier (Université de Montréal) 97

Introduction 97
1. Les instruments de mesure 98
1.1 Les objectifs du test 98
1.2 Le contenu du test 100
1.3 La mise au point 101
2. La comparaison entre les versions 103
2.1 Première étude 104
2.1.1 Méthodologie 104
2.1.2 Résultats 105
2.1.3 Discussion 110
2.2 Deuxième étude 110
2.2.1 Méthodologie 110
2.2.2 Le test adaptatif 111
2.2.3 Le test conventionnel 113
2.2.4 Discussion 114
Conclusion 115

SIXIÈME CHAPITRE – ÉVALUER LA TRANSFORMATION DE STRUCTURES DE CONNAISSANCE PROPOSITIONNELLE CHEZ DES ADULTES EN FORMATION : QUESTIONS MÉTHODOLOGIQUES POUR LA RECHERCHE
Étienne Bourgeois (Université catholique de Louvain) 119

Introduction 119
1. Une perspective constructiviste de l'apprentissage 120
2. Implications méthodologiques du paradigme constructiviste 124
2.1 Accès aux connaissances par les représentations circonstancielles 124
2.2 Accès aux connaissances par inférence 125
3. Implications pour le recueil des données 126
3.1 Premier type de recherche 126
3.2 Deuxième type de recherche 129
4. Implications pour le traitement des données 131
Conclusion 132

SEPTIÈME CHAPITRE – CROYANCES SAILLANTES ET CONDITIONS DE PRATICABILITÉ DE L'ÉVALUATION FORMATIVE DES APPRENTISSAGES TELLES QUE PERÇUES PAR LES ENSEIGNANTS
Serge P. Séguin, Carmen Parent, et Rose Burelle 137
(Université du Québec à Montréal)

Introduction 137
1. Rappel du modèle 137
2. Rappel de l'instrument 138
3. Croyances saillantes 140
4. Échantillons 141
5. Résultats 142
5.1 L'intention 143
5.2 L'attitude 144
5.3 Les croyances normatives 147
5.4 La motivation 147
5.5 Les croyances comportementales 149
6. Croyances saillantes 154
Conclusion 156

HUITIÈME CHAPITRE – L'ACTE D'ÉVALUER LES APPRENTISSAGES : PROBLÉMATIQUE DE LA FORMATION
Louise Bélair (Université d'Ottawa) 161

Introduction 161
1. L'acte d'évaluer au primaire : une action humaine intériorisée 162
1.1 L'acte d'évaluer implique un jugement moral complexe 162
1.2 L'acte d'évaluer est différent selon l'évaluateur 164
1.3 L'acte d'évaluer est différent selon l'objet 166
1.4 L'acte d'évaluer est différent selon l'évalué 167
1.5 L'acte d'évaluer est avant tout interactif 169
1.6 Une réalité ... troublante! 169
1.7 La formation actuelle au Québec et en Ontario 171
2. Évaluer les apprentissages des élèves : une question d'apprentissage 172
2.1 L'évaluation perçue encore comme une innovation 173
2.2 Apprendre à évaluer, c'est comme apprendre à lire! 174
2.3 Projets de formation et changement planifié 176
Conclusion 179

NEUVIÈME CHAPITRE – LA PLACE PRIMORDIALE À LA PERSONNE DE L'ÉDUCATEUR POUR LE CHANGEMENT DE SA PRATIQUE ÉVALUATIVE
Richard Robillard (Université Sainte-Anne) 185

Introduction 185
1. Rétrospective de mes expériences d'intervenant en évaluation 192
2. De la recherche d'un acte idéal d'évaluation des apprentissages à la pratique évaluative «réelle» de l'éducateur 198
2.1 La recherche d'un acte idéal d'évaluation des apprentissages 199
2.2 L'émergence d'une série de mouvements 201
3. Faire une place à la personne de l'éducateur 204
3.1 Que signifie «faire une place à la personne de l'éducateur»? 204
3.2 Qu'est-ce qui nous motive à «faire une place primordiale à la personne de l'éducateur»? 209
3.3 Comment faire une place à la personne de l'éducateur pour le changement de sa pratique évaluative? 218
3.4 Le processus de «prise de contact» avec l'univers de sa personne 219
3.5 Processus de «prise de parole» 220
Conclusion 222

DIXIÈME CHAPITRE – ÉTUDE DE LA VALIDITÉ DES ÉPREUVES DE FRANÇAIS ÉCRIT DE SIXIÈME ANNÉE DU PRIMAIRE DE JUIN 1990 ET DE JUIN 1991 EN CONFORMITÉ AVEC LE PROGRAMME DE FRANÇAIS LANGUE MATERNELLE DU MINISTÈRE DE L'ÉDUCATION DU QUÉBEC
Réjean Auger (Université du Québec à Montréal) 229

Introduction 229
1. Problématique 230
2. Cadre d'analyse 231
3. Méthodologie 234
4. Analyse 236
Conclusion 242

PRÉSENTATION

Ce volume présente les contributions des chercheurs qui ont participé au symposium du Réseau de recherche en éducation et formation qui s'est tenu à Sherbrooke, en octobre 1993, sous le thème spécifique «L'évaluation des apprentissages». Les auteurs traitent de cette dimension dans le processus enseignement et apprentissage.

Dans une première partie, des auteurs aborderont la question de l'évaluation des apprentissages à la lumière des recherches récentes en psychologie cognitive et en évaluation formative. Par la suite, d'autres discuteront de l'utilisation de l'informatique en évaluation formative et sommative. Suivront quatre textes qui toucheront l'éducateur et l'évaluation : croyances, formations, pratiques. À la toute fin, un texte important sur le concept de validité dans le contexte des examens du ministère de l'Éducation.

Le premier chapitre «Objectifs terminaux d'intégration et transfert des connaissances» par Jean-Marie De Ketele présente une description des pratiques les plus courantes, une synthèse des résultats de la recherche qui permet d'évaluer les pratiques évaluatives en terme de transfert de connaissance. Ce texte permet de retrouver l'évolution des objets d'évaluation.

Le deuxième chapitre «L'évaluation dans le paradigme constructiviste» par Jacques Tardif identifie les contributions récentes de la recherche. Après cette revue, l'auteur présente le concept «évaluation authentique» et comment se situe l'évaluation dans le cadre de l'enseignement de stratégiques.

Le troisième chapitre «L'évaluation formative du processus d'apprentissage : le rôle des régulations métacognitives» par Linda Allal discute des recherches portant sur l'implication de l'élève dans le processus de l'évaluation formative, soit l'autorégulation de l'élève. L'importance de la métacognition dans l'apprentissage et dans l'évaluation est discutée.

Le quatrième chapitre «Contexte de mesure et évaluation informative» par Clément Dassa définit et propose une approche instrumentée de l'évaluation formative qui permettrait d'intégrer l'évaluation formative à l'enseignement réel et à l'apprentissage vécu en classe. L'apport des nouvelles technologies dans l'enseignement et son utilité dans un contexte d'évaluation formative est nettement mis en évidence.

Le cinquième chapitre «Un test adaptatif en langue seconde : la perce[illegible]nants» par Michel Laurier discute des effets de la médiat[illegible]quement de l'informatisation de l'évaluation et de la percepti[illegible]comparativement à l'approche conventionnelle.

Le sixième chapitre «Évaluer la transformation de structure de connaissance propositionnelle chez les adultes en formation» présenté par Étienne Bourgeois cherche à répondre à la question «savoir comment on peut identifier et évaluer les transformations cognitives qui s'opèrent chez les adultes en formation?».

Le septième chapitre «Croyances saillantes et conditions de praticabilité de l'évaluation formative des apprentissages telles que perçues par les enseignants» par Serge P. Séguin, Carmen Parent, Rose Burelle a pour objectif de contribuer à la compréhension des raisons qui motivent les enseignants à adopter ou non diverses pratiques d'évaluation formative dans leurs classes. Cette étude contribue à mieux situer les enseignants et leurs perceptions face à l'évaluation formative.

Le huitième chapitre «L'acte d'évaluer les apprentissages : la problématique de la formation» par Louise Bélair veut amener le lecteur ou la lectrice à se positionner face à sa propre démarche d'évaluation et à la démarche de formation continue qu'il ou qu'elle entreprend.

Le neuvième chapitre «La place primordiale à la personne de l'éducateur pour le changement de sa pratique éducative» par Richard Robillard présente les réflexions de l'auteur face à cet acte professionnel et accorde principalement l'importance à la personne de l'éducateur. L'auteur nous fait découvrir son cheminement personnel et ses propres réflexions sur les pratiques évaluatives.

Le dixième chapitre «Études de la validité des épreuves de français écrit, de sixième année du primaire de juin 1990 et juin 1991 en conformité avec le programme de français langue maternelle du ministère de l'Éducation du Québec» par Réjean Auger présente une étude sur la validité des épreuves en titre. Les résultats obtenus permettent à l'auteur de conclure que les pratiques évaluatives du Ministère sont parcellaires, car elles intègrent difficilement le processus d'évaluation fondé sur le paradigme information-jugement décision.

L'évaluation n'est pas une fin en soi. Elle se doit d'être un processus qui permet à l'enseignant et à l'élève de prendre des décisions appropriées en leur fournissant l'information pour les prendre et intervenir. Chaque coauteur de cet ouvrage collectif tente de répondre, sous des facettes particulières, aux questions relatives à la nature de ce qui est mesuré et évalué, aux qualités nécessaires de la mesure, aux croyances des élèves et des enseignants et spécifiquement à la place de la cognition et de la métacognition. L'évaluation «authentique» est-elle une réponse? En définitive, comment s'exerce ce processus? Pouvons-nous faire mieux?

René Hivon
Université de Sherbrooke

OBJECTIFS TERMINAUX D'INTÉGRATION ET TRANSFERT DES CONNAISSANCES

JEAN-MARIE DE KETELE
Laboratoire de pédagogie expérimentale
Université catholique de Louvain

INTRODUCTION

«Dites-moi comment vous évaluez et je vous dirai ce que vos élèves ou vos étudiants apprennent réellement!» Nous pourrions, dans bien des cas, ajouter «et je vous dirai votre conception réelle de l'apprentissage».

En accord avec cet adage, notre communication repose sur trois temps : a) une description des pratiques les plus courantes d'évaluation des apprentissages; b) une synthèse de résultats de recherches; c) cette synthèse permet d'évaluer les pratiques évaluatives en termes de transfert de connaissances.

1. INVENTAIRE DESCRIPTIF DES PRATIQUES ÉVALUATIVES

1.1 Les pratiques centrées sur les contenus

Cette pratique, la plus courante, consiste à prélever un échantillon de contenus représentatifs de l'univers de référence des contenus enseignés. Chaque contenu peut être décomposé en un univers fini d'unités de contenu plus fines. Chacune de celles-ci peut faire l'objet d'un ensemble (fini ou infini, selon les cas) de questions potentielles parmi lesquelles seront choisies une ou plusieurs questions selon la longueur souhaitée de l'évaluation.

Dans cette optique, la pertinence de l'évaluation dépendra essentiellement de la valeur de l'univers de référence. La validité de l'épreuve d'évaluation se caractérise par le caractère représentatif de l'échantillon par rapport à l'univers de référence, par la position de la moyenne (pas trop éloignée du centre de l'échelle de notation), par la dispersion des résultats (variance la plus grande possible pour discriminer au mieux les élèves) et par la forme de la distribution (courbe gaussienne).

Ces exigences expliquent certaines opérations de transformation des résultats originaux observés lorsqu'ils n'obéissent pas aux caractéristiques de celle-ci le long de l'échelle autour d'une moyenne plus «vraisemblable», procédure de «modération» dans les examens nationaux ou régionaux (pratique fréquente en Grande-Bretagne)...

1.2 Les pratiques centrées sur les activités

Un autre point d'entrée, plus naturel chez certains enseignants, réside dans les activités susceptibles d'être générées par une matière donnée. Dans cette optique, il existerait un ensemble d'activités potentielles de référence liées à une matière, parmi lesquelles l'évaluateur choisirait un échantillon (se réduisant parfois à une seule activité) d'activités. Dans ce cas, nous retombons dans les schémas précédents, si ce n'est que l'unité de l'évaluation est l'activité au lieu d'être le contenu ou l'objectif spécifique.

1.3 Les pratiques centrées sur les objectifs spécifiques

Ces pratiques, issues du mouvement de la technique des objectifs pédagogiques (TOP), supposent au départ que soit établi l'univers de référence en termes d'«arbre d'objectifs» : on se fixe des objectifs généraux et on les décompose en termes d'objectifs de plus en plus spécifiques et opérationnels.

L'évaluation consiste alors à prendre un échantillon représentatif d'objectifs spécifiques et opérationnels et à générer un échantillon de questions qui traduise au mieux cet échantillon d'objectifs. On est ainsi ramené aux règles mentionnées précédemment. Cependant, les critères de correction ne sont pas centrés sur la quantité (éventuellement pondérée) d'informations pertinentes présentes dans la réponse de l'élève ou de l'étudiant, mais sur la présence ou l'absence d'indicateurs (éventuellement pondérés) censés être le signe de la maîtrise des objectifs sélectionnés dans l'échantillon.

1.4 Les pratiques centrées sur les réseaux conceptuels

Issues d'un réseau de recherches didactiques, ces pratiques – il est vrai très rares, car elles supposent une maîtrise conceptuelle approfondie des disciplines rarement atteinte explicitement par les experts eux-

mêmes – supposent que le domaine objet de l'apprentissage est essentiellement de l'ordre du savoir déclaratif et que celui-ci doit pouvoir être exprimé sous forme d'un ensemble de concepts définis de façon précise, en relation et hiérarchisés.

À partir de là, plusieurs conceptions de l'évaluation peuvent être générées; on peut a) échantillonner un ensemble de couples de concepts et examiner dans quelle mesure les concepts et leurs relations sont compris de façon précise (le nombre de relations entre concepts augmente très vite avec l'étendue de la matière); b) présenter à l'étudiant un réseau conceptuel incomplet et lui faire combler les lacunes (il existe un nombre élevé de réseaux conceptuels lacunaires incomplets); c) interroger l'étudiant à partir du concept le plus englobant pour tester la compréhension (au sens étymologique du terme, «prendre ensemble les éléments pour former un tout»).

Les deux premières stratégies obéissent aux mêmes règles que précédemment, car il s'agit de prélever un échantillon représentatif de questions pour discriminer les élèves ou les étudiants. La troisième stratégie est d'un autre ordre, car un seul concept «intègre» les autres.

1.5 Les pratiques centrées sur une activité ou thème intégrateur

Assez fréquentes chez les partisans d'une «éducation nouvelle» et chez certains innovateurs, ces pratiques consistent à choisir une activité ou un thème susceptibles de mettre en jeu les contenus ou/et activités ou/et thèmes les plus importants qui ont fait l'objet des apprentissages antérieurs. Il ne s'agit donc pas de choisir un échantillon d'activités ou de thèmes, mais bien d'imaginer une activité ou un thème intégrateur susceptibles d'intéresser les élèves. L'évaluateur sera ainsi amené à faire un bilan des contenus et des comportements censés avoir été appris antérieurement et judicieusement réutilisés.

Ces bilans peuvent être faits, soit dans une optique normative : les étudiants sont situés les uns par rapport aux autres (rangement en fonction de la quantité éventuellement pondérée de contenus ou de comportements réutilisés), soit dans une optique critériée : des critères de maîtrise, minimaux et de perfectionnement, sont déterminés sans se préoccuper d'une norme de classement.

1.6 Les pratiques centrées sur les objectifs d'intégration

Nous définissons un objectif d'intégration comme une compétence qui, dans l'idéal, possède les quatre caractéristiques suivantes. En premier lieu, la compétence s'exerce sur une situation d'intégration, c'est-à-dire sur une situation complexe qui comprend tant de l'information essentielle que parasite et qui met en jeu les apprentissages antérieurs: c'est une caractéristique minimale nécessaire. En deuxième lieu, la compétence est une activité complexe qui nécessite l'intégration et non la juxtaposition de savoirs et de savoir-faire antérieurs fondamentaux et qui aboutit à un produit évaluable qui les intègre. En troisième lieu, la situation d'intégration est la plus proche possible d'une situation naturelle que pourra rencontrer l'élève : elle a donc une fonction sociale. En quatrième lieu, l'objectif d'intégration fera appel à des savoir-être et à des savoir-devenir orientés vers le développement de l'autonomie. Voici deux exemples qui illustrent ce concept d'objectif terminal d'intégration.

Au terme de l'enseignement primaire, on peut évaluer les acquis en mathématiques en demandant aux élèves de faire un devis des travaux (compétence et non simple juxtaposition de capacités) à effectuer dans l'école, un plan de celle-ci, une liste de prix hors et avec TVA étant fournis (situation d'intégration). En effet, tous les acquis fondamentaux en mathématique peuvent être mis en œuvre à partir de cette compétence à exercer sur cette situation.

Pour un cours de méthode élémentaire de recherche, on peut demander aux étudiants de rédiger le rapport scientifique conformément à des normes reconnues (compétence) d'une recherche qu'ils auront menée selon certaines indications données (situation d'intégration).

Deux grandes catégories de pratiques évaluatives peuvent être étudiées conformément à ce paradigme : la compétence évaluée et la situation d'intégration sur laquelle elle s'exerce sont imposées à l'étudiant; celles-ci sont laissées pour une part plus ou moins grande à l'étudiant.

Il existe un très grand nombre de compétences et de situations d'intégration potentielle à propos d'un cours déterminé. La pertinence de l'évaluation dépend avant tout de la détermination des savoirs et des

savoir-faire (et même des savoir-être et des savoir-devenir), considérés comme fondamentaux, impliqués dans la compétence à évaluer. Il faut, en outre, noter que la démarche d'intégration doit être objet d'apprentissage progressif.

2. LES DÉTERMINANTS DE LA PERFORMANCE ACADÉMIQUE

Pour jeter un regard critique sur ces diverses pratiques évaluatives, il semble intéressant au préalable de synthétiser les principaux résultats des très nombreuses recherches internationales qui ont essayé de mettre en évidence les déterminants de la performance académique (Voir à ce sujet Parmentier, De Ketele *et al.*, 1993). Ceux-ci devraient nous permettre de mettre en lumière le degré de pertinence des pratiques évaluatives.

La plupart des recherches s'entendent pour dire qu'il n'existe guère de relation entre la quantité de contenus-matières vus dans les années antérieures et la performance académique ultérieure. Par contre, il existerait une relation étroite (estimée à environ 25 % de la variance expliquée, si on tient constantes les autres variables explicatives) entre les démarches cognitives exercées sur les contenus, quels qu'ils soient, et la performance académique. Ces démarches cognitives, que nous appelons capacités cognitives de base, ne sont pas nombreuses et elles commencent à être bien connues. Citons, à titre d'exemples, la capacité à traduire en langage formel ce qui a été exprimé en langage de tous les jours (et réciproquement), plus globalement la maîtrise de la langue d'apprentissage, distinguer l'essentiel de l'accessoire, prendre des notes utiles, structurer un message, résumer, synthétiser, etc.

La maîtrise de ces démarches cognitives représente une condition nécessaire mais non suffisante de la performance académique. Sur ce fondement viennent se greffer les variables plus affectives et conatives : l'intérêt pour les disciplines de base d'un programme d'étude, l'effort consenti, le caractère significatif des apprentissages, l'image de soi et surtout la perception de l'image renvoyée par certaines autres personnes concernant la capacité à réussir les études entreprises... Ces facteurs liés aux premiers expliqueraient environ 45 % de la variance de la performance académique.

Sur ces deux fondements viendraient se greffer deux autres séries de facteurs. D'une part, le processus de choix d'une filière d'étude et, d'autre part, la façon de gérer sa vie d'étudiant; ceci implique de nombreuses variables (capacité à gérer son temps, à équilibrer les différentes activités, à adopter des stratégies d'études adéquates, etc.).

L'ensemble de ces facteurs permettrait de rendre compte d'environ 80 % de la variance de la performance scolaire, du moins au niveau de l'enseignement supérieur et universitaire.

À côté de ces recherches sur ce qui détermine la réussite scolaire, il est intéressant de mentionner les recherches qui tentent d'établir le lien entre les performances académiques et la réussite dans le milieu professionnel, évaluée sur la base d'indicateurs tant objectifs (par exemple, le niveau de salaire, le nombre de subordonnés, ...) que subjectifs (degré d'épanouissement perçu, désir de rester ou de changer d'établissement). Ici encore, les recherches semblent indiquer une absence de relation entre les résultats évalués par les enseignants, le plus souvent en termes de matières vues, et la réussite professionnelle.

3. ÉVALUATION DES PRATIQUES ÉVALUATIVES

L'évaluation des effets des pratiques évaluatives suppose que l'on se pose le problème crucial des critères de référence. En ce qui nous concerne et sur la base des résultats de recherches précédentes, nous prendrons en considération les critères suivants : a) le degré de profondeur des acquis les plus vraisemblables générés par la conception de l'apprentissage sous-jacente aux pratiques évaluatives; b) la conservation des acquis, à court terme et à long terme; c) le degré d'intérêt suscité à plus ou moins long terme; d) le degré d'intériorisation des effets des apprentissages évalués; e) le degré de transfert; f) le degré d'intégration des acquis.

À la lecture de ces critères, on s'aperçoit aisément qu'ils ne sont pas strictement indépendants. Ainsi, par exemple, des acquis bien intériorisés ont plus de chance de se conserver à long terme.

3.1 Les pratiques centrées sur les contenus

Des recherches qui ont fait une analyse de contenu des pratiques évaluatives centrées sur les contenus montrent qu'environ 95 % des

questions d'évaluation sont de l'ordre du savoir-redire ou du savoir-refaire ce qui a été dit ou fait lors de l'enseignement. Et nous savons, par ailleurs, que ce type de connaissances résiste très mal au temps.

On peut douter que soit grand l'intérêt généré par une conception de l'apprentissage et de l'évaluation où «le tout est considéré comme la somme des parties». L'évaluation sommative porte particulièrement bien son nom. On peut cependant imaginer que des enseignants enthousiastes puissent créer un intérêt pour leur champ disciplinaire. Mais ce n'est pas de loin une règle générale, comme le suggère, par exemple, l'enquête française sur l'intérêt pour les sciences : celui-ci, très vif au départ, décroît rapidement avec l'introduction d'un enseignement des sciences plus classique (plus formel et centré sur les contenus inscrits au programme) dans la scolarité.

Les pratiques évaluatives majoritairement centrées sur le savoir-redire et le savoir-refaire incitent davantage les élèves et les étudiants à adopter des «approches d'étude superficielles» ou des «approches d'étude stratégiques» (au sens défini par des auteurs comme Entwistle, 1983, 1990)). Il y a donc peu d'intériorisation des connaissances soumises à l'étude, d'autant plus que celles-ci ne sont guère mises en relation avec le vécu des apprenants.

Dans cette conception de l'apprentissage, le transfert est quasiment nul : seul ce qui est très proche de ce qui a été dit ou fait peut être maîtrisé.

Cette conception de l'apprentissage est donc une pédagogie de l'accumulation de connaissances juxtaposées, une «pédagogie à tiroirs» : on ouvre et on referme successivement des tiroirs de contenus-matières. C'est donc tout le contraire d'une pédagogie de l'intégration.

3.2 Les pratiques centrées sur les activités

Les pratiques centrées sur les activités privilégient les applications qui rendent actifs les élèves et leur permettent d'exploiter les connaissances théoriques du champ disciplinaire. À ce titre, elles regroupent surtout des catégories taxinomiques que Gagné (1970) qualifiait d'«application algorithmique» (mise en jeu d'un algorithme qu'on vient d'appren-

dre) et d'«application heuristique», ce qui implique le choix d'un ou de plus d'un algorithme parmi un ensemble d'algorithmes connus.

Cette pédagogie centrée sur l'activité de l'élève est davantage motivante pour l'élève, dans la mesure où celui-ci reçoit suffisamment de renforcements positifs par la réussite des tâches entreprises.

Il n'est cependant pas sûr que l'élève intériorise ce qui semble acquis : trop souvent, les activités sont faites pour elles-mêmes et sont insuffisamment reliées les unes aux autres, condition strictement requise pour intérioriser et pour conserver les acquis.

Dans cet ordre d'idée, on peut alors rarement parler de transfert, car seules les activités réussies peuvent l'être à nouveau. Pour cela, il faudrait mener des apprentissages progressifs à l'intégration.

3.3 Les pratiques centrées sur les objectifs spécifiques

Les pratiques évaluatives centrées sur les objectifs spécifiques ont le mérite de savoir ce qu'elles évaluent et d'orienter les apprentissages en conséquence. L'apprentissage est donc rarement centré exclusivement sur des savoir-redire ou refaire.

La communication des objectifs aux élèves et la conscience par ceux-ci de capacités maîtrisées de plus en plus nombreuses peuvent être motivantes. Trop souvent cependant, ces objectifs apparaissent trop formels et sans signification pour les élèves. Leur accumulation dans de longues listes accroît ce sentiment.

La juxtaposition d'objectifs, tant sur le plan de l'apprentissage que sur celui de l'évaluation, explique sans doute la médiocrité du niveau de conservation des acquis et du niveau de transfert constatés tant par les praticiens que par les chercheurs.

Avec ce modèle, on assiste à une autre forme de «pédagogie à tiroirs», très béhavioriste dans la conception et où, même dans les objectifs de haut niveau taxinomique, l'intégration est peu présente.

3.4 Les pratiques centrées sur les réseaux conceptuels

La conception de l'apprentissage sous-jacente à ces pratiques met l'accent sur les relations qui existent entre les concepts qui fondent un champ disciplinaire. Elle forme des esprits rigoureux, qui pourront apprendre d'autres connaissances plus aisément, car les fondations conceptuelles sont solides. Ces pratiques garantissent donc davantage la conservation des acquis.

Ces pratiques conviendront bien aux esprits rationnels qui aiment la formalisation. Elles risquent cependant de peser pour les esprits, et ils sont nombreux, qui préfèrent les approches concrètes.

Un réseau de concepts correctement articulés est favorable au transfert, du moins en ce qui concerne les activités formelles. Cependant, l'absence de va-et-vient entre théorie et pratique, très fréquent dans les apprentissages de ce type, nuit à une véritable intériorisation des acquis.

Il s'agit bien d'une pédagogie de recherche du sens formel, mais pas nécessairement de recherche de la signification *in situ*. Il n'y a donc pas, dans ce cas, de véritable intégration.

3.5 Les pratiques centrées sur une activité ou thème intégrateur

Ce type de pratique a comme mérite majeur d'impliquer les élèves, du moins dans la mesure où les activités ou thèmes n'ont pas été choisis sur la seule base des intérêts de l'enseignant.

Ce qui a été agi par l'élève dans un climat de motivation et avec à la clé des résultats jugés positifs se conserve plus aisément à long terme.

Quant à dire quelles sont les capacités acquises à travers les activités exercées, les enseignants s'en préoccupent souvent très peu ou ont de la difficulté à les déterminer. On peut cependant estimer qu'elles dépassent le simple niveau des savoir-redire ou des savoir-refaire.

Dans la mesure où le lien explicite est fait entre contenus du programme à voir et les activités exercées sur des situations d'intégration, les contenus revêtent un sens pour les élèves. On peut donc espérer que

certains transferts à d'autres situations pas trop éloignées puissent être effectués; ceci requiert cependant que les élèves aient pu travailler sur des situations suffisamment nombreuses et variées. Pour optimaliser davantage le transfert, il faudrait davantage travailler les relations formelles entre concepts fondamentaux; c'est loin d'être une préoccupation courante de ces pratiques.

Dans ce type de pratiques, il semble que les élèves ou étudiants intériorisent et intègrent des démarches et des outils méthodologiques beaucoup plus que des concepts; il n'est d'ailleurs pas impossible que des activités de nature intégratrice soient correctement et efficacement menées alors que, fondamentalement, les apprenants ont conservé des conceptions erronées concernant des concepts de base.

3.6 Les pratiques centrées sur les objectifs d'intégration

Ce type de pratiques privilégie non pas une juxtaposition de capacités ou d'objectifs spécifiques, mais la mise en œuvre de compétences, explicitées cette fois-ci, qui s'exercent sur des situations qui ont du sens pour l'élève.

Elles installent donc des acquis qui dépassent de loin les simples savoir-redire ou savoir-refaire et qui assurent une très bonne conservation des acquis. Ce faisant, ces activités sont le plus souvent motivantes, par le caractère actif des apprentissages et par le caractère significatif des situations.

Dans la mesure où l'apprentissage a su doser les objectifs d'intégration progressivement et a travaillé sur des situations suffisamment nombreuses et variées, on peut s'attendre à un plus haut niveau de transfert que dans les autres conceptions.

Contrairement au modèle précédent, le danger n'existe pas, en principe du moins, de ne privilégier que l'acquisition des outils méthodologiques. En effet, la conception d'un objectif terminal d'intégration suppose que l'on ait déterminé au préalable les concepts et les objectifs spécifiques les plus importants des apprentissages antérieurs. Ainsi donc, les processus d'intériorisation et d'intégration devraient être d'autant plus importants qu'en outre, les situations d'apprentissage sont significatives et progressives pour l'apprenant.

CONCLUSION

Nous sommes bien conscients que tout apprentissage s'exerce sur des contenus, implique plus ou moins explicitement des objectifs spécifiques, initie des activités, repose sur des réseaux conceptuels plus ou moins mis en évidence. C'est peut-être l'intégration qui manque le plus dans les activités d'apprentissage que nous avons observées. C'est cependant, à nos yeux, une composante tout à fait nécessaire pour assurer un transfert de haut niveau et pour voir les acquis se conserver à long terme. De plus, une pédagogie de l'intégration est plus propice à créer un véritable intérêt.

Mais si la recherche pédagogique est en mesure d'établir des liens entre situations d'enseignement d'une part et effets de processus d'intériorisation et d'intégration d'autre part, elle reste jusqu'à présent assez impuissante à mettre en évidence les processus eux-mêmes. Les travaux sur l'évaluation des apprentissages ne progresseront significativement que si des avancées dans ce domaine sont réalisées.

RÉFÉRENCES

ENTWISTLE, N. et HILARY, T. (1990).
Approaches to learning, evolutions of teaching, and references for contrasting academic environments, *Higher Education, 19* 169-194.

ENTWISTLE, N. et RAMSDEN, J. (1983).
Understanding student learning. London : Croom Helm.

GAGNÉ, R.M. (1970).
The conditions of learning. (New York : Holt, Rinehart and Winston.

PARMENTIER, P., DE KETELE, J.-M. *et al.* (1993).
Recherches et initiatives dans le domaine de la pédagogie universitaire à l'Université catholique de Louvain : bilan et impact sur la performance académique en candidatures (1987-1992). Document interne, Université catholique de Louvain.

L'ÉVALUATION DANS LE PARADIGME CONSTRUCTIVISTE

JACQUES TARDIF
Université de Sherbrooke

INTRODUCTION

Peu importe l'ordre d'enseignement pris en considération, les pratiques évaluatives qui ont présentement cours dans les salles de classe et les institutions scolaires soulèvent un très grand nombre d'interrogations. Pour les enseignants en particulier et pour la société en général, ces pratiques évaluatives rendent-elles compte des métamorphoses cognitives de l'élève? Les évaluations, autant formatives que sommatives, ont-elles pour objet la connaissance au détriment de la compréhension? Le contenu qui est retenu dans les démarches et dans les instruments d'évaluation fait-il partie des connaissances antérieures de l'élève? Quels sont 'es liens que ce contenu entretient avec la base de connaissances spécifiques de l'élève? La structure même des instruments d'évaluation donne-t-elle accès à la construction du savoir? Cette structure est-elle plus orientée vers l'évaluation des habiletés cognitives de bas niveau que vers l'évaluation des habiletés cognitives de haut niveau? Le contexte dans lequel se déroule l'évaluation introduit-il des variables affectives qui influencent négativement le rendement de l'élève? Les instruments permettent-ils de rendre compte systématiquement des forces et des faiblesses de l'élève? Les critères de correction respectent-ils la valeur de chacune des réponses de l'élève au regard de la construction d'un savoir particulier? Ces critères sont-ils influencés davantage par des exigences d'objectivité que par la valeur cognitive des réponses?

Cette multitude de questions – d'autres pourraient être ajoutées – souligne à la fois la complexité et l'importance de la problématique de l'évaluation dans le milieu scolaire et dans le cheminement cognitif de l'élève. Les questions mentionnées concernent autant le contenu des instruments d'évaluation, leur structure, leurs objectifs, le contexte dans lequel prend place l'évaluation que les critères de correction. La problématique de l'évaluation est toutefois difficile à circonscrire sans faire

explicitement référence à la conception de l'apprentissage qui est à la base des modalités et des instruments retenus aux fins d'évaluation. Les pratiques évaluatives témoignent toujours d'une conception spécifique de la nature de l'apprentissage et du développement de l'expertise cognitive, bien que cette conception soit fréquemment implicite.

Il importe cependant de souligner que la conception de l'apprentissage exerce des influences non seulement sur les pratiques évaluatives, mais également sur les pratiques d'enseignement. Il y a beaucoup de cohérence entre la conception qu'un enseignant a de l'apprentissage et ce qu'il met en place pour l'enseignement et l'évaluation. Il en est d'ailleurs ainsi pour les divers agents dans le milieu scolaire. Selon l'optique de ces interactions, Perrone (1991) signale que, si l'enseignement demeure une question de feuilles et de cahiers d'exercice, de manuels de base et d'explications simplifiées, il est impossible que les pratiques évaluatives soient différentes de ce qu'elles sont maintenant. Dans un contexte de réflexion et de propositions au sujet de l'évaluation, il devient alors capital de prendre directement en considération les principes se référant à l'apprentissage et d'en déduire les conséquences sur les pratiques d'enseignement et sur les pratiques d'évaluation.

À la base de cette présentation, je poursuis l'objectif de développer l'idée que seules les tâches complètes, complexes et signifiantes pour l'élève peuvent constituer une situation valide d'évaluation sur le plan cognitif. Cette idée est en conformité avec les objectifs terminaux de l'école qui, pour la très grande majorité, voire la totalité, sont de l'ordre des compétences. Les compétences, quelles qu'elles soient, ne peuvent être circonscrites dans un court laps de temps. Elles ne sont pas réductibles à un certain nombre d'items précis, de la même façon qu'elles ne sont pas identifiables à partir d'un ou de quelques types de questions. Enfin, elles ne peuvent être observées à partir de questions qui portent exclusivement sur quelques composantes d'une compétence donnée. Dans les écrits de langue anglaise, en faisant référence à ces tâches complètes, complexes et signifiantes, les auteurs insistent fréquemment sur le caractère «authentique» de l'évaluation et, selon cette perspective, ils ont introduit le concept d'évaluation authentique. Ce concept est capital dans les nouvelles pratiques évaluatives.

Un tel objectif exige d'abord une définition de ce qu'est l'évaluation dans le but de la distinguer de la mesure et du *testing* et dans le but de rendre explicites ses frontières et ses caractéristiques. Par la suite, cet objectif impose que les principes d'apprentissage selon divers paradigmes soient présentés et que leurs conséquences sur les pratiques d'enseignement et d'évaluation soient discutées. À ce moment-là, les conséquences relatives aux principes constructivistes sont l'objet d'une discussion plus approfondie. Dans la poursuite de l'objectif à la base de ce chapitre, les caractéristiques de l'évaluation authentique sont mentionnées et quelques-unes sont décrites. Enfin, les critères de validité de l'évaluation en conformité avec la conception constructiviste de l'apprentissage sont discutés brièvement.

1. UNE DÉFINITION DE L'ÉVALUATION

Dans le *Dictionnaire actuel de l'éducation*, Legendre (1988) définit ainsi l'évaluation des apprentissages : «Appréciation des connaissances, des habiletés ou des attitudes d'un élève qui a pour objet de déterminer dans quelle mesure des objectifs éducatifs ont été atteints» (p. 259). Cette définition, même si elle est très importante pour les institutions scolaires, demeure exclusivement axée sur l'atteinte d'objectifs éducatifs au détriment du cheminement cognitif de l'élève, au détriment de ses métamorphoses cognitives et métacognitives. Par l'évaluation, les institutions scolaires ont, entre autres, la préoccupation de rendre compte publiquement des objectifs atteints grâce à la scolarisation et la définition de Legendre (1988) est très logique dans ce sens. Dans les écrits de langue anglaise, les auteurs insistent alors sur le rôle public (*accountability*) de l'évaluation.

La définition de Legendre (1988) est insatisfaisante parce qu'elle ne met pas l'accent sur les métamorphoses cognitives et métacognitives de l'élève. Elle n'est centrée que sur l'atteinte d'objectifs éducatifs sans faire référence explicitement au cheminement cognitif de l'élève. Pour l'enseignant et l'élève, comme pour les institutions scolaires, la connaissance des métamorphoses de l'élève est plus importante que la connaissance du nombre d'objectifs atteints et de leur niveau d'atteinte à la fin de telle ou telle étape, de telle ou telle année, de tel ou de tel ordre d'enseignement. Selon Zessoules et Gardner (1991), les pratiques évaluatives

doivent permettre une description exacte des métamorphoses cognitives de l'élève; elles doivent fournir des informations précises sur ses forces et sur ses faiblesses.

Gardner (1992) définit l'évaluation comme le processus par lequel des informations sont obtenues au sujet des connaissances (*skills*) et des capacités (*potentials*) d'une personne et il ajoute, dans la même phrase, que ce processus de recueil d'informations vise à fournir une rétroaction significative à la personne évaluée et à fournir des données utiles à la communauté dans laquelle elle vit. C'est cette conception de l'évaluation qui est retenue dans la présente communication, et cela pour plusieurs raisons.

Obtenir des informations «au sujet des connaissances et des capacités d'une personne» (Gardner, 1992) dépasse énormément le fait d'apprécier des connaissances, des habiletés ou des attitudes «dans le but de déterminer dans quelle mesure des objectifs éducatifs ont été atteints» (Legendre, 1988). Dans le dernier cas, il s'agit essentiellement d'indiquer le degré d'atteinte d'objectifs particuliers alors que, dans le premier, il s'agit de décrire, de fournir des informations sur les métamorphoses de l'élève. Dans ces distinctions, il n'y a pas que des mots, il y a une conception de l'évaluation. Dans l'esprit de la définition de Gardner (1992), l'évaluation doit mettre l'accent sur les connaissances que l'élève a développées, sur les étapes de ces développements et sur les capacités de l'élève. L'évaluation privilégie la base de connaissances que l'élève a intégrées en mémoire à long terme au lieu d'être centrée sur des manques et des lacunes ou encore sur des objectifs atteints ou non.

Selon Zessoules et Gardner (1991), l'évaluation prend place à la fin et au début d'une démarche d'apprentissage : au début, pour déterminer les connaissances antérieures de l'élève; à la fin, pour déterminer ce qu'il a appris. L'évaluation est fondamentalement une activité située sur les démarches d'apprentissage et la définition de Gardner (1992) est également très cohérente à cet égard. C'est justement parce que l'évaluation tient compte des connaissances antérieures de l'élève et des connaissances acquises dans une démarche d'apprentissage qu'il est possible de décrire ses métamorphoses cognitives et métacognitives.

La définition de Gardner (1992), à l'instar de plusieurs autres cependant, a le mérite d'établir une nette distinction entre l'évaluation, la mesure et le *testing*. La mesure est une activité qui vise exclusivement «la description quantitative des connaissances, des capacités ou des habiletés d'un élève» (Legendre, 1988, 366). L'évaluation peut donner lieu à des conclusions quantitatives, mais telle n'est pas sa visée première. Conceptuellement, il n'y a pas de contraintes pour que les retombées de l'évaluation soient quantitatives. L'évaluation est une appréciation, un recueil d'informations et, dans ce sens, les données et les conclusions peuvent être autant qualitatives que quantitatives.

Dans les discussions sur les pratiques évaluatives, le concept de *testing* est fréquemment mis en relation ou opposé à celui d'évaluation. Selon Legendre (1988, 609), le *testing* est une «opération de mesure ou d'évaluation qui prend place avant ou après une situation pédagogique [...]». Il ajoute aussi que c'est l'action d'administrer un test. Le *testing* est une forme d'évaluation ou de mesure parmi plusieurs autres et l'évaluation comme la mesure ne sauraient être réduites au *testing*. Cependant, il importe de reconnaître que, dans les pratiques évaluatives actuelles, l'administration de tests à des fins d'évaluation occupe une telle place qu'il est dangereux de confondre l'évaluation et les tests. Cette confusion serait dramatique puisque, selon Perrone (1991), le *testing* est le mode d'évaluation le plus faible et le plus ambigu au regard des connaissances et des capacités d'un élève.

Dans la définition qu'il présente au sujet de l'évaluation, Gardner (1992) inclut deux buts, soit celui de fournir une rétroaction significative à la personne évaluée et celui de fournir des données utiles à la communauté dans laquelle vit cette personne. Il y a plusieurs classifications des buts de l'évaluation dans le cadre scolaire. Lesgold, Lajoie, Logan et Eggan (1990) indiquent que l'évaluation poursuit cinq buts : 1) décrire les compétences d'un élève; 2) expliquer le niveau de performance atteint par un élève; 3) permettre l'adaptation des interventions d'enseignement aux compétences réelles d'un élève; 4) déterminer les difficultés d'un élève pour une compétence donnée; 5) contribuer à la sélection des élèves. Pour leur part, Resnick et Resnick (1992) signalent les trois buts suivants : 1) évaluer un programme et rendre publiques les retombées des

pratiques d'enseignement; 2) déterminer les actions d'enseignement et le soutien pédagogique nécessaires; 3) sélectionner les élèves et attester d'un niveau d'études.

Dans le cadre de notre réflexion présente, la prise en considération des buts reconnus à l'évaluation ne vise pas à discuter de leur pertinence ou de leur vraisemblance, ni du caractère exhaustif des classifications proposées. Il s'agit plutôt d'insister sur le fait que des auteurs, entre autres Resnick et Resnick (1992), estiment qu'une même évaluation ou que des modalités identiques d'évaluation ne peuvent prétendre répondre à plusieurs buts. Selon eux, il faudrait des formes et des modalités d'évaluation spécifiques pour chacun des buts. En privilégiant la définition de Gardner (1992), définition dans laquelle deux buts généraux sont retenus et dans laquelle l'évaluation est conçue comme le recueil d'informations sur des connaissances et des capacités de l'élève, il semble qu'une même évaluation, une même forme d'évaluation authentique toutefois, puisse répondre à plusieurs buts, voire à l'ensemble des buts d'évaluation dans le cadre scolaire.

Par conséquent, l'évaluation authentique est proposée comme susceptible de fournir une réponse très adéquate et très appropriée pour l'ensemble des buts poursuivis dans les situations d'évaluation à l'école. Frederiksen (1990) affirme que la nécessité des évaluations sommatives finales pourrait diminuer dans la mesure où s'amélioreront les moyens d'extraire des informations pertinentes sur les connaissances lors du processus même d'apprentissage. De très grandes responsabilités professionnelles seraient alors octroyées à l'enseignant.

2. QUELQUES PRINCIPES D'APPRENTISSAGE

Dans le domaine de l'enseignement et de l'apprentissage, on reconnaît que trois paradigmes psychologiques ont eu des retombées extrêmement importantes : le paradigme associationniste, le paradigme humaniste et le paradigme constructiviste. Étant donné l'état des connaissances et des recherches au regard de l'apprentissage, les principes issus du paradigme constructiviste sont les plus susceptibles de provoquer et d'assister l'acquisition d'un savoir significatif, réutilisable fonctionnellement et transférable. Selon le paradigme constructiviste, l'apprentissage est fondamentalement une activité de traitement d'informations et les

principes relatifs au traitement de l'information sont essentiellement ceux qui correspondent à la nature de l'apprentissage. Lors du traitement d'informations à des fins d'apprentissage, l'élève, comme toute personne, ne peut photocopier l'ensemble des informations qui sont mises à sa disposition. Il doit sélectionner les plus importantes pour en extraire la connaissance, pour construire le savoir. C'est entre autres dans cette perspective qu'il est question d'un paradigme constructiviste. Il est aussi question de constructivisme dans l'optique où l'élève, lors du traitement d'informations, construit des lois, des règles ou des principes auxquels il octroie rapidement un caractère de permanence dans sa base de connaissances en mémoire à long terme.

Avant de décrire les principes d'apprentissage reliés au paradigme constructiviste, il importe de souligner que la psychologie cognitive s'inscrit dans ce paradigme et que c'est à la lumière de données de la psychologie cognitive que ces principes seront abordés. Il importe également de souligner que, en enseignement et en apprentissage, le paradigme constructiviste s'est développé en tenant compte des retombées cognitives insatisfaisantes qui résultent des démarches et des scénarios prescrits par les paradigmes humaniste et associationniste. Dans le but de rendre explicites les influences de ces deux paradigmes et, surtout, de mieux situer les principes constructivistes, les tableaux 1 et 2 présentent rapidement leurs conséquences sur l'enseignement, sur l'apprentissage et sur l'évaluation. Dans le tableau 1, comme dans la suite du texte, le béhaviorisme est considéré comme la théorie psychologique associationniste et les conséquences mentionnées pour le paradigme associationniste sont issues de cette théorie.

Selon le paradigme associationniste, les divers éléments se réunissent par association automatique. En enseignement et en apprentissage, ce principe a été appliqué intégralement et toutes les connaissances, quelles qu'elles soient, ont été considérées comme se développant par association automatique. Ce principe d'association a conduit à la hiérarchisation des connaissances et à la détermination de préalables. Certaines connaissances sont nécessairement préalables à d'autres, certains répertoires de comportements sont antérieurs à d'autres. Dans cette orientation, le contenu présenté à l'élève est morcelé et ce dernier agit constamment sur les composantes d'un ensemble plus large. Le

morcellement du contenu est une conséquence majeure du paradigme associationniste dans les pratiques d'enseignement et d'évaluation.

Tableau 1 - Conséquences du paradigme associationniste sur l'enseignement, sur l'apprentissage et sur l'évaluation

1. Conception de l'enseignement
   - 1.1 Création d'un environnement qui contraint l'association stimulus-réponse.
   - 1.2 Création d'un environnement qui est axé sur le développement de comportements.
   - 1.3 Création d'un environnement qui morcelle le contenu.
   - 1.4 Création d'un environnement qui organise le contenu d'un préalable à un autre.
   - 1.5 Création d'un environnement coercitif par l'enseignant.
2. Conception de l'apprentissage
   - 2.1 L'apprentissage s'effectue par association stimulus-réponse.
   - 2.2 L'apprentissage se fait essentiellement par imitation.
   - 2.3 L'apprentissage se fait par approximations successives.
3. Conception de l'évaluation
   - 3.1 L'évaluation est très fréquente.
   - 3.2 L'évaluation concerne uniquement les comportements produits.
   - 3.3 L'évaluation porte sur des petites unités.
   - 3.4 L'évaluation est souvent formative, parfois sommative.
   - 3.5 L'évaluation fournit une mesure.
   - 3.6 La rétroaction concerne le produit de la performance.
4. Conception de l'apprenant
   - 4.1 L'apprenant répond aux stimuli de l'environnement.
   - 4.2 L'apprenant est réactif.
   - 4.3 L'apprenant a une motivation contrôlée par des renforcements extérieurs.

Tableau repris et adapté de Tardif (1992)

La conséquence la plus importante de ce paradigme en enseignement, en apprentissage et en évaluation a toutefois été de conduire l'enseignant à ne prendre en considération que la performance de l'élève, le produit de ses démarches. Le produit est directement observable et mesurable selon les béhavioristes et il s'agit là de la seule donnée fiable dans les interventions humaines. Appliqué dans l'enseignement et dans l'évaluation, ce principe a fait en sorte que l'enseignant mette l'accent sur la performance et néglige, voire ignore, les processus qui assurent l'atteinte de tel ou tel produit. Dans le paradigme associationniste, le produit est retenu au détriment du processus.

Selon le paradigme humaniste, paradigme moins influent dans le cadre de l'enseignement et de l'apprentissage que le paradigme associationniste, l'apprentissage se fait par l'exploration et par la découverte. L'enseignant peut assister l'élève dans ses propres démarches, mais il ne saurait intervenir directement dans les motivations personnelles de ce dernier et dans ses besoins de croissance. Le rôle crucial de l'enseignant est de créer un environnement susceptible de participer à la croissance de l'élève. L'enseignant est essentiellement un animateur qui met en place un environnement de tâches complètes et signifiantes. Cet aspect relatif à la signifiance des activités est capital. Il faut surtout noter que le paradigme humaniste s'oppose fortement à l'évaluation formelle et sommative. L'observation directe de l'élève lors de la réalisation de tâches permet d'estimer avec satisfaction ses apprentissages et, surtout, sa croissance personnelle et son cheminement.

Tableau 2 – Conséquences du paradigme humaniste sur l'enseignement, sur l'apprentissage et sur l'évaluation

1. Conception de l'enseignement
   - 1.1 Création d'un environnement riche et très varié.
   - 1.2 Création d'un environnement axé sur la croissance personnelle.
   - 1.3 Création d'un environnement à partir des besoins de l'élève.
   - 1.4 Création d'un environnement de tâches complètes et signifiantes.
   - 1.5 Création d'un environnement libre.
2. Conception de l'apprentissage
   - 2.1 L'apprentissage se fait par découverte personnelle.
   - 2.2 L'apprentissage se fait par découverte collective.
   - 2.3 L'apprentissage se fait par exploration.
   - 2.4 L'apprentissage s'effectue à partir d'un contenu global.
3. Conception de l'évaluation
   - 3.1 L'évaluation formelle est inexistante.
   - 3.2 La rétroaction concerne essentiellement la croissance personnelle.
4. Conception de l'apprenant
   - 4.1 L'apprenant est actif.
   - 4.2 L'apprenant choisit lui-même les activités.
   - 4.3 L'apprenant a une motivation intrinsèque issue de ses besoins de croissance personnelle.

Tableau repris et adapté de Tardif (1992)

Il a été mentionné que, dans l'enseignement et l'apprentissage, le paradigme constructiviste s'est développé en tenant compte des retombées cognitives insatisfaisantes qui résultent des démarches et des scénarios prescrits par les paradigmes humaniste et associationniste. Entre autres, ce paradigme s'oppose à la hiérarchisation et au morcellement du contenu typique du paradigme associationniste et il estime, contrairement au paradigme humaniste, que le savoir se construit, dans la salle de classe, grâce aux interventions directes, explicites et fréquentes de l'enseignant. Le tableau 3 présente les cinq principes qui caractérisent l'apprentissage dans le paradigme constructiviste.

Tableau 3 – Principes de base du paradigme constructiviste au regard de l'apprentissage

- L'apprentissage est un processus actif et constructif.
- L'apprentissage est l'établissement de liens entre de nouvelles informations et les connaissances antérieures.
- L'apprentissage requiert l'organisation constante des connaissances.
- L'apprentissage concerne tout autant les stratégies cognitives et métacognitives que les connaissances théoriques.
- L'apprentissage concerne autant les connaissances déclaratives, procédurales que conditionnelles.

Tableau repris de Tardif (1992)

Le premier principe concerne le fait que l'apprentissage est un processus actif et constructif. En situation d'apprentissage, l'élève agit directement sur les informations présentées. Il sélectionne quelques informations parmi celles qui sont mises en sa présence et ce processus de sélection exige l'utilisation de stratégies qui permettent non seulement d'interagir significativement avec ces informations, mais également de rejeter celles qui sont secondaires ou de peu d'importance. Les informations ne s'associent pas entre elles automatiquement. L'élève doit mettre en opération un processus de sélection et, à cet égard, le principe d'association des informations l'une à l'autre pour qu'elles se transforment en connaissances ne parvient pas à expliquer la réalité de l'apprentissage. Plusieurs recherches, particulièrement des recherches sur la représentation de la connaissance chez l'élève (Giordan et de Vecchi,

1987) ont contribué à illustrer le caractère nécessairement constructif de l'apprentissage. Dans ces recherches, le concept du champ de validité de la représentation de la connaissance est d'ailleurs capital.

Le fait que l'apprentissage soit l'établissement de liens entre de nouvelles informations et les connaissances antérieures s'avère le second principe du paradigme constructiviste. Ce principe n'est toutefois pas nouveau puisqu'il avait été énoncé par Ausubel en 1968. Ces dernières années, il a été souvent repris, plus spécifiquement par Jones, Palincsar, Ogle et Carr (1987). À ce principe s'ajoute, en corollaire, l'idée que l'apprentissage est essentiellement un processus cumulatif dans le sens où les nouvelles connaissances sont reliées aux connaissances antérieures. Selon ce principe, l'élève doit avoir en mémoire à long terme des connaissances qui lui permettent d'établir des liens avec les nouvelles informations qui lui sont présentées, sinon il ne peut les traiter d'une façon significative. À l'égard de ce principe, il est capital de mentionner que les connaissances antérieures exercent un rôle à ce point prédominant dans l'apprentissage qu'elles résistent souvent aux informations introduites par l'enseignant quand ces dernières vont à l'encontre des connaissances antérieures.

Le troisième principe du paradigme constructiviste au regard de l'apprentissage concerne le fait que l'apprentissage requiert l'organisation constante des connaissances. Dans la mémoire à long terme de l'expert, les connaissances sont organisées et hiérarchisées, et cette organisation ainsi que cette hiérarchisation des connaissances sont responsables de l'expertise cognitive développée. Dans le paradigme constructiviste, l'organisation des connaissances en mémoire est reconnue comme un trait caractéristique de l'expertise. Le fait que les connaissances soient organisées et hiérarchisées en mémoire à long terme augmente les probabilités que la personne intègre de nouvelles informations d'une façon significative et qu'elle puisse réutiliser fonctionnellement ses connaissances. Les recherches de Meyer (Meyer, 1975; Meyer, Brandt et Bluth, 1980; Meyer et Rice, 1984; Meyer, 1985; Meyer, 1987) qui portent sur les effets de la connaissance des structures textuelles en compréhension de texte illustrent très bien le rôle et les effets de l'organisation des connaissances en mémoire.

Un quatrième principe de base du paradigme constructiviste a trait au fait que l'apprentissage concerne tout autant les stratégies cognitives et métacognitives que les connaissances théoriques. C'est plus particulièrement selon cette optique que de nombreux auteurs distinguent la connaissance de la compréhension. Le système cognitif de la personne ne contient pas seulement des connaissances factuelles ou statiques, il contient des connaissances «pratiques» ou dynamiques. Les connaissances théoriques sont réputées être des connaissances statiques. Elles permettent de nommer le réel et de l'objectiver, mais elles ne permettent pas d'agir sur lui. C'est par le biais de connaissances dynamiques, surtout grâce aux stratégies cognitives et métacognitives, que la personne peut agir sur le réel. D'une part, les stratégies cognitives fournissent à l'élève les moyens d'utiliser fonctionnellement les connaissances acquises, de poser les actions adéquates dans les diverses tâches qu'il est appelé à réaliser. D'autre part, les stratégies métacognitives lui permettent de gérer activement les stratégies qu'il utilise pour réaliser ces tâches.

Le cinquième principe du paradigme constructiviste indique que l'apprentissage concerne autant les connaissances déclaratives que les connaissances conditionnelles et procédurales. La détermination de ces catégories de connaissances résulte, dans un premier temps, des travaux d'Anderson (1983, 1985) et, dans un deuxième temps, de distinctions introduites plus particulièrement par Marzano, Brandt, Hugues, Jones, Presseisen, Rankin et Suhor (1988). La distinction de ces catégories de connaissances est extrêmement importante puisque leur représentation spécifique en mémoire à long terme commande des stratégies d'enseignement fort différentes. Les connaissances déclaratives sont essentiellement des connaissances théoriques. Il s'agit, selon Gagné (1985), de la connaissance de faits, de règles, de lois, de principes. Les connaissances procédurales correspondent au comment de l'action, aux étapes pour réaliser une action; il s'agit de séquences d'actions. Les connaissances conditionnelles concernent le quand et le pourquoi des actions. Les connaissances conditionnelles et procédurales sont aussi importantes dans le système cognitif de la personne que les connaissances déclaratives et l'enseignement doit agir sur toutes ces catégories de connaissances pour que l'élève puisse apprendre et agir significativement.

Le tableau 4 présente les conséquences du paradigme constructiviste sur l'enseignement, l'apprentissage et l'évaluation. Les conséquences au regard de l'évaluation sont considérées attentivement à la suite de ce tableau.

Tableau 4 – Conséquences du paradigme constructiviste sur l'enseignement, l'apprentissage et l'évaluation

1. Conception de l'enseignement
   - 1.1 Création d'un environnement à partir des connaissances antérieures de l'élève.
   - 1.2 Création d'un environnement axé sur les stratégies cognitives et métacognitives.
   - 1.3 Création d'un environnement axé sur l'organisation des connaissances.
   - 1.4 Création d'un environnement de tâches complètes, complexes et signifiantes.
   - 1.5 Création d'un environnement coercitif.
2. Conception de l'apprentissage
   - 2.1 L'apprentissage se fait par construction graduelle des connaissances.
   - 2.2 L'apprentissage se fait par la mise en relation des connaissances antérieures et des nouvelles informations.
   - 2.3 L'apprentissage exige l'organisation des connaissances.
   - 2.4 L'apprentissage s'effectue à partir de tâches globales.
3. Conception de l'évaluation
   - 3.1 L'évaluation est fréquente.
   - 3.2 L'évaluation porte sur les compétences.
   - 3.3 L'évaluation concerne les connaissances ainsi que les stratégies cognitives et métacognitives.
   - 3.4 L'évaluation est souvent formative, parfois sommative.
   - 3.5 La rétroaction est axée sur les stratégies utilisées.
   - 3.6 La rétroaction est axée sur la construction du savoir.
4. Conception de l'apprenant
   - 4.1 L'apprenant est actif.
   - 4.2 L'apprenant est constructif.
   - 4.3 L'apprenant a une motivation en partie déterminée par la perception de la valeur de la tâche et du contrôle qu'il peut avoir sur sa réussite.

Tableau repris et adapté de Tardif (1992)

Comme il est possible de le constater à la lecture du tableau 4, les retombées du paradigme constructiviste sur les pratiques évaluatives sont nombreuses et diversifiées. Une première conséquence extrêmement importante concerne les instruments mêmes d'évaluation. Dans le paradigme constructiviste, contrairement aux prescriptions du paradigme

associationniste, les apprentissages doivent être réalisés à partir de tâches complètes, complexes et signifiantes. En conformité avec les pratiques d'enseignement, les évaluations reposent nécessairement sur des tâches complètes, complexes et signifiantes. Elles ne sauraient, conceptuellement et opérationnellement, privilégier des composantes au détriment des compétences. C'est spécifiquement dans cette perspective qu'il est fréquemment mentionné que la compréhension de l'élève et non ses connaissances doit être privilégiée lors des évaluations. Ici cependant, les connaissances sont entendues dans le sens restrictif de connaissances déclaratives. Dans cet esprit, en faisant référence aux tests utilisés actuellement dans le milieu scolaire, Wiggins (1992) déclare qu'ils sous-évaluent la compréhension de l'élève et qu'ils surévaluent ses connaissances théoriques.

Une autre conséquence du paradigme constructiviste sur les pratiques évaluatives, conséquence très cohérente avec celle susmentionnée, a trait à la prise en compte, directement et explicitement, des stratégies cognitives utilisées par l'élève. Ces stratégies ne peuvent être évaluées à partir de simples questions qui portent sur les composantes d'une compétence. Les situations qui peuvent révéler quelque chose au regard des stratégies cognitives et métacognitives des élèves sont essentiellement des tâches complètes, complexes et signifiantes. Celles-ci permettent de déterminer avec exactitude l'usage que l'élève peut faire de ses connaissances déclaratives, conditionnelles et procédurales ainsi que leur degré de transférabilité. Cette conséquence insiste encore une fois sur la nécessité pour l'évaluation de dépasser la connaissance pour rejoindre la compréhension. Elle permet également de souligner, dans la logique des conclusions de Linn, Baker et Dunbar (1991), que toute évaluation exige une analyse de la complexité cognitive des tâches proposées ainsi qu'une analyse de la nature des réponses que ces tâches sont susceptibles de provoquer.

Il a été mentionné que la personne ne peut traiter significativement que les informations qui ont des liens avec ses connaissances antérieures, avec ses connaissances en mémoire à long terme. Ce principe fondamental dans l'apprentissage a des retombées cruciales pour les pratiques évaluatives. En effet, le contenu retenu lors des démarches

d'évaluation doit être assez familier pour l'élève, de sorte que ce dernier puisse établir des liens entre ce contenu et ses connaissances en mémoire à long terme. Toute évaluation qui ne garantit pas ce genre de liens ne peut permettre à l'élève de manifester les connaissances déclaratives, procédurales et conditionnelles qu'il a développées. L'analyse de la complexité cognitive des tâches proposées à des fins d'évaluation exige que les connaissances antérieures de l'élève soient considérées très attentivement. Le débat au sujet de la prépondérance des connaissances spécifiques ou des connaissances générales prend, dans ce contexte, une grande signification.

Le système cognitif de la personne inclut des connaissances déclaratives, des connaissances conditionnelles et des connaissances procédurales et l'école, dans ses objectifs terminaux, ambitionne de contribuer au développement de ces diverses catégories de connaissances. Toute évaluation doit exiger de l'élève, en conformité avec les objectifs d'enseignement, le recours à des connaissances déclaratives, conditionnelles et procédurales. Ces deux dernières catégories de connaissances permettent l'action sur le réel et il est capital qu'elles soient incluses dans l'ensemble des démarches évaluatives. Les connaissances conditionnelles sont particulièrement importantes parce que, selon les données issues de la psychologie cognitive, elles sont à la base même du transfert des apprentissages. Dans le milieu scolaire, puisque les évaluations doivent rendre compte de la construction du système cognitif de l'élève et de son évolution, il est nécessaire que toutes les catégories de connaissances soient l'objet de chaque démarche évaluative.

Le troisième principe du paradigme constructiviste au regard de l'apprentissage concerne le fait que l'apprentissage requiert l'organisation constante des connaissances. L'organisation et la hiérarchisation des connaissances, comme cela a été mentionné précédemment, est un trait distinctif de l'expertise cognitive. En accord avec ce principe, les pratiques d'enseignement selon le paradigme constructiviste font en sorte que l'enseignant intervient directement et fréquemment sur l'organisation et sur la hiérarchisation des connaissances. C'est entre autres ainsi que l'enseignant participe étroitement à la construction du savoir par l'élève. Les évaluations doivent systématiquement rendre compte de

l'état de l'organisation et de la hiérarchisation des connaissances par l'élève. Toute pratique évaluative doit permettre de tirer des conclusions, avec exactitude, sur la progression de l'élève dans la construction du savoir et cette progression est nécessairement caractérisée par l'organisation et la hiérarchisation des connaissances.

Dans le paradigme constructiviste, le savoir est une construction graduelle. D'une part, l'enseignement assiste et provoque cette construction graduelle. D'autre part, l'évaluation rend compte de l'évolution de cette construction. Selon cette optique, il importe que les pratiques évaluatives dépassent le constat statique précisant un «degré donné de connaissances» pour décrire l'évolution cognitive de l'élève. Dans cet esprit, Lidz (1987), à la suite de Feurstein (1969), souligne qu'il est nécessaire que l'evaluation renseigne sur les changements cognitifs survenus, sur les domaines dans lesquels ces changements se manifestent et, puisque Lidz (1987) privilégie un cadre de médiation, elle demande que l'évaluation renseigne également sur le nombre d'interventions requises pour provoquer ces changements ainsi que sur leur nature. L'évaluation est une démarche qui vise essentiellement à décrire les métamorphoses cognitives de l'élève.

Les stratégies cognitives et métacognitives sont de première importance dans les prescriptions du paradigme constructiviste au sujet de l'enseignement et de l'apprentissage. Dans la perspective de la construction du savoir, les stratégies métacognitives méritent toutefois une attention particulière. Par le biais de ces stratégies, l'élève gère activement les démarches qu'il réalise. Les stratégies métacognitives sont autant présentes lors de la phase de planification d'une activité que lors des phases de réalisation et d'évaluation de cette activité. Elles sont des habiletés cognitives de haut niveau et elles concernent la conscience et le contrôle de ce que la personne met en action pour réaliser une démarche. Le caractère crucial des stratégies métacognitives tient au fait qu'elles garantissent à la personne une très grande autonomie dans la réalisation de ses démarches. Leur importance est, entre autres, illustrée par les conclusions de plusieurs recherches qui indiquent que ces stratégies distinguent les experts des novices et les élèves qui réussissent de ceux qui sont en difficulté. Il est donc essentiel que les évaluations intègrent

des situations qui prennent directement en considération les stratégies métacognitives utilisées par l'élève. À cet égard, la nécessité de tâches complètes, complexes et signifiantes lors des évaluations est encore soulignée. L'évaluation authentique est de première importance.

3. QUELQUES CARACTÉRISTIQUES DE L'ÉVALUATION AUTHENTIQUE

Resnick et Resnick (1992) signalent que, à la base des tests standardisés, il y a deux postulats majeurs. Ces postulats sont d'ailleurs à la base de l'ensemble des tests traditionnels. Ils nomment le premier le postulat de la «décomposabilité» et celui-ci correspond à l'idée que toute compétence[1] est décomposable en composantes et que l'évaluation de chaque composante, isolément, fournit des indices significatifs pour ce qui est de la compétence. Le deuxième postulat est celui de la décontextualisation. Il implique que chaque composante d'une compétence est fixe et qu'elle se manifeste de la même façon peu importe le contexte dans lequel elle prend place, peu importe le contenu à partir duquel elle est évaluée.

Resnick et Resnick (1992) mentionnent que ces deux postulats ont été sérieusement mis en doute par les recherches inscrites dans les orientations du paradigme constructiviste. Dans la mesure où, pour chacune des compétences, les interactions entre leurs composantes beaucoup plus que les composantes elles-mêmes sont les déterminants du niveau de développement d'une compétence donnée, la «décomposabilité» est un postulat invraisemblable. Ce postulat conduit à privilégier l'évaluation de chaque composante isolément alors que ce sont les interactions entre les composantes qui sont les variables les plus lourdes et les plus significatives dans le développement de toute compétence.

En ce qui concerne la décontextualisation, les recherches portant sur le rôle des connaissances spécifiques dans la compétence ont particulièrement contribué à rejeter ce postulat. Un très grand nombre de recherches ont démontré que plus une personne a de connaissances spécifiques sur un sujet donné, plus sa performance dans ce domaine est élevée. La

---

1 Le terme anglais *performance*, étant donné le sens qui lui est attribué par la majorité des auteurs de langue anglaise dans le contexte de l'évaluation authentique, est traduit par le mot «compétence» dans le présent texte.

décontextualisation ignore systématiquement le rôle capital des connaissances spécifiques dans la manifestation d'une compétence. Le postulat de la décontextualisation a ainsi contribué à ce que les instruments utilisés dans les phases d'évaluation fassent référence à des contenus peu signifiants et à ce que le transfert des connaissances soit traité d'une façon réductionniste. À l'égard du manque de signifiance des instruments utilisés, une conclusion de Gardner (1992) mérite d'être reprise. Il souligne que la platitude est l'une des caractéristiques inacceptables des tests formels.

Ce qui est rapporté de Resnick et Resnick (1992) est en étroite relation avec les principes du paradigme constructiviste au regard de l'enseignement et de l'apprentissage et avec les conséquences de ces principes sur les pratiques évaluatives. Les deux postulats susmentionnés ne sauraient être à la base des pratiques évaluatives dans le paradigme constructiviste. Pellegrino (1992) insiste sur le fait que l'évaluation et les tests doivent être en conformité avec les théories qui concernent la nature de l'expertise et de l'apprentissage. Dans ce contexte, les seules évaluations qui respectent les principes du paradigme constructiviste sont celles qui privilégient directement la compétence contextualisée. Il y aurait alors deux nouveaux postulats à la base des pratiques évaluatives dans le cadre du paradigme constructiviste. Le premier est celui de l'interactivité des composantes dans les compétences. Le deuxième est celui de la contextualisation des compétences.

L'évaluation qui tient compte de la contextualisation de la compétence et des interactions de ses composantes exige des tâches complètes, complexes et signifiantes. La signifiance de la tâche est cruciale étant donné la très grande importance des connaissances spécifiques dans la manifestation d'une compétence. De plus, la tâche doit être complète parce qu'il faut qu'elle oblige le recours aux interactions entre les composantes de la compétence évaluée. Comme il a été mentionné, ce sont les interactions entre les composantes d'une compétence beaucoup plus que le nombre de composantes qui permettent de conclure avec exactitude sur le degré de maîtrise atteint pour toute compétence. Enfin, la tâche est complexe parce que la situation d'évaluation doit placer l'élève dans un contexte où l'usage de stratégies cognitives et métacognitives est nécessaire.

Les évaluations qui respectent ces caractéristiques ne correspondent certainement pas à des activités qui peuvent se réaliser dans un court laps de temps. Elles ne peuvent, d'ailleurs, être constituées que de questions fermées ou de tableaux réductionnistes à compléter. Le tableau 5 signale les caractéristiques que Wiggins (1989) reconnaît aux évaluations authentiques. Ces informations sont présentées à titre indicatif et elles ne sont pas reprises telles quelles dans le texte. Au lieu de les reprendre, j'ai choisi de discuter des porte-folios qui illustrent très bien une démarche d'évaluation authentique.

Tableau 5 - Caractéristiques des évaluations authentiques

Pour ce qui est de la conception des modalités d'évaluation

1. La production de l'élève est destinée à un jury.
2. Il n'y a aucune contrainte de temps fixée arbitrairement.
3. La tâche et ses exigences sont connues avant la situation d'évaluation.
4. L'évaluation partage les caractéristiques des porte-folios.
5. L'évaluation exige une certaine forme de collaboration avec des pairs.
6. Les tâches de l'évaluation authentique sont récurrentes.
7. L'évaluation et la rétroaction sont à ce point importantes que l'horaire de la classe est fait pour rendre ces activités les plus signifiantes possible pour l'élève.

Pour ce qui est de l'orientation cognitive des tâches

1. L'évaluation est essentielle dans la construction du savoir.
2. L'évaluation contribue à ce que l'élève développe encore davantage ses compétences.
3. L'évaluation n'inclut que des tâches contextualisées.
4. L'évaluation exige l'utilisation fonctionnelle de connaissances spécifiques.
5. L'évaluation prend en considération les stratégies cognitives et métacognitives utilisées par l'élève.
6. L'évaluation présente des défis visant le traitement en profondeur des informations par l'élève.
7. L'évaluation inclut des problèmes mal définis.
8. L'évaluation demande beaucoup d'engagement de la part de l'élève.

Pour ce qui est des critères de correction

1. La correction ne tient compte que des erreurs importantes dans l'optique de la construction du savoir.
2. Les critères de correction sont déterminés en faisant référence aux exigences cognitives de la compétence.
3. Les critères de réussite sont démystifiés pour l'élève.
4. L'auto-évaluation fait partie de l'évaluation.
5. Les critères de correction sont multiples et donnent lieu à plusieurs informations sur la compétence évaluée.
6. Les critères de correction sont en conformité avec les buts de l'école.

Pour ce qui est de l'équité et de la déontologie

1. L'évaluation doit déterminer les forces de l'élève.
2. Les informations extraites de l'évaluation doivent établir un équilibre rigoureux entre les aptitudes de l'élève, ses connaissances antérieures et les nouveaux apprentissages réalisés.
3. L'évaluation doit minimiser les comparaisons inutiles, injustes et démoralisantes entre les élèves.
4. L'évaluation doit laisser une place importante aux différences individuelles ainsi qu'aux intérêts de l'élève.
5. Les mêmes démarches d'évaluation sont exigées de tous les élèves et le soutien nécessaire est disponible pour ceux qui éprouvent des difficultés.
6. L'évaluation est guidée par les exigences de la validité écologique et elle est avant tout au service de l'apprentissage.

Tableau traduit et adapté de Wiggins (1989)

Dans le milieu de l'enseignement aux États-Unis, Wiggins a fortement influencé les discussions sur les pratiques évaluatives et il a contribué à introduire le concept d'évaluation authentique. Selon Meyer (1992), toute évaluation authentique est nécessairement une évaluation d'une compétence. Les problèmes réels à résoudre, les essais, les projets de recherche, les simulations, les activités de laboratoires, etc. sont des situations d'évaluation authentique. Les situations d'évaluation authentique doivent permettre, le plus directement possible, de tirer des conclusions au sujet des compétences développées par l'élève. Ces situations visent à décrire et à certifier ce que l'élève est comme penseur, lecteur, scripteur, mathématicien, etc. L'évaluation authentique, selon les propos de Resnick et Resnick (1992), tient compte directement des compétences plutôt que de les considérer à l'aide d'indicateurs indirects.

Depuis quelques années, les pratiques reliées à l'authenticité de l'évaluation connaissent un essor considérable. Toutefois, Perrone (1991) fait remarquer que l'évaluation authentique en milieu scolaire a une longue histoire. Il souligne que ce qui est appelé documentation, porte-folio, évaluation de la compétence et «exposition» des apprentissages, correspond à des pratiques qui étaient courantes dans les écoles progressives du XIX^{e} siècle, écoles influencées par John Dewey, William Kilpatrick, Marietta Johnson et Caroline Pratt. De plus, depuis plusieurs années, l'Angleterre et la Nouvelle-Zélande recourent à des pratiques d'évaluation authentique. C'est reconnaître que les pratiques

d'évaluation authentique ne sont pas nécessairement récentes dans les classes et les institutions scolaires, et que des conclusions fort pertinentes peuvent être extraites des expériences déjà réalisées ou en cours.

Particulièrement aux États-Unis, les porte-folios constituent un moyen de plus en plus fréquemment mis en place pour répondre aux exigences de l'évaluation authentique et déterminer les métamorphoses cognitives de l'élève. À titre d'exemple, O'Neil (1992) cite le Kentucky qui, en 1996, aura un programme d'évaluation principalement axé sur les compétences. Dès 1992, les élèves de quatrième, de huitième et de douzième année ont cependant utilisé les porte-folios en écriture et l'évaluation a porté sur ces documents. O'Neil (1992) mentionne également le Vermont où, en 1991, en mathématique et en écriture, les élèves de quatrième et de huitième année ont employé les porte-folios. L'évaluation formelle de l'État du Vermont était basée sur ces porte-folios. Pour O'Neil (1992), l'urgence et la nécessité d'évaluer les élèves pour ce qui est du développement des habiletés cognitives de haut niveau, comme la capacité de résoudre des problèmes complexes, de travailler en collaboration, d'intégrer des informations provenant de champs disciplinaires diversifiés conduit plusieurs états vers des pratiques d'évaluation authentique.

Les porte-folios constituent un moyen très efficace d'accéder aux compétences maîtrisées par l'élève et de déterminer leur degré de développement. Dans le cadre des porte-folios, les responsabilités accordées à l'élève dans son cheminement cognitif sont nombreuses et elles contribuent directement au développement de la gestion de ses connaissances, au développement de stratégies métacognitives. De plus, l'usage des porte-folios à des fins d'évaluation fait en sorte que les enseignants assument de nouvelles responsabilités qui exigent le développement de compétences professionnelles essentielles pour assister la construction du savoir. Toutes ces raisons militent en faveur des modalités d'évaluation relatives aux porte-folios ou de modalités semblables.

Les porte-folios, comme toute pratique d'évaluation authentique, poursuivent le but de fournir un portrait le plus complet possible des capacités et des métamorphoses cognitives de l'élève. La démarche la

plus fréquemment utilisée est la suivante : 1) pour une période donnée (une étape, une année, un cycle, un ordre), l'élève est informé qu'une partie de ses productions sera utilisée à des fins d'évaluation; 2) pour la période déterminée, l'élève sélectionne les productions qui, selon lui, illustrent le mieux ses compétences ainsi que leur évolution; 3) l'échantillon des productions retenues est soumis à un jury dont la composition est variable selon les finalités de l'évaluation en question. Les porte-folios exigent donc que l'élève évalue lui-même, dans un premier temps, la valeur de représentation de ses productions en terme de construction du savoir et qu'il en sélectionne quelques-unes en conséquence. C'est l'élève qui détermine les productions qui sont retenues pour l'évaluation.

Le recours aux porte-folios est aussi très exigeant pour l'enseignant. Il faut qu'il précise les critères d'évaluation qui permettent de bien situer les compétences maîtrisées par l'élève ainsi que leur degré de développement. Selon Resnick et Resnick (1992), la décision d'adopter des évaluations axées sur la compétence exige que les enseignants définissent très précisément ce qui est désirable sur le plan cognitif et il faut que ces critères d'excellence soient déterminés pour chacune des compétences en question. Dans la même perspective, Pellegrino (1992) ajoute que les critères doivent être spécifiés à partir d'une théorie de la compétence et que les démarches évaluatives doivent être cohérentes au regard de cette théorie de la compétence. Toujours dans la perspective des critères que l'enseignant doit adopter pour les évaluations authentiques, Brown, Campione, Webber et McGilly (1992) insistent sur la nécessité d'analyser le contenu de chaque champ de connaissances, de circonscrire le cheminement cognitif dans ce champ de connaissances et préciser les problèmes que l'élève est susceptible de rencontrer dans ce domaine.

Cependant préciser des critères d'évaluation dans le cadre des porte-folios est d'autant plus complexe et difficile que la culture de l'évaluation qui a prévalu lors de la période associationniste a placé une très grande importance sur l'objectivité, sur la validité et sur la fidélité des instruments de mesure. Cette culture marque encore grandement les réflexions et les décisions relatives aux instruments et aux modalités d'évaluation. Les critères de fidélité et de validité privilégiés dans le

paradigme constructiviste sont fort différents de ceux privilégiés dans le paradigme associationniste et la détermination de ces critères est encore l'objet de discussion sur le plan conceptuel alors que les enseignants ont la contrainte d'agir sur le plan opérationnel.

4. AU SUJET DE LA FIDÉLITÉ ET DE LA VALIDITÉ

En ce qui concerne les instruments d'évaluation dans le paradigme associationniste, le concept de fidélité occupe une place de première importance. Selon Linn, Baker et Dunbar (1991), la fidélité a reçu une attention démesurée alors que la validité a été traitée d'une façon réductionniste. Ces auteurs insistent toutefois sur la nécessité de travailler simultanément les critères d'adéquation des instruments d'évaluation et les formes d'évaluation. Et, dans la même orientation que Messick (1989), ils attribuent une place privilégiée à la validité de conséquence ou la validité déontologique. Avant d'aborder les critères de validité des instruments dans le cadre du paradigme constructiviste, il importe toutefois de considérer la fidélité.

Au sujet de la fidélité, Chittenden (1991) estime que ce critère relatif aux instruments est crucial pour toute évaluation qui est réalisée en une seule période de temps et dont les retombées pour la personne évaluée sont très importantes. Il poursuit en précisant que ce critère est secondaire dans la mesure où les conclusions au regard des compétences d'un élève sont issues de plusieurs situations et qu'il s'agit d'un échantillon des productions de l'élève en question. Dans cette optique, la fidélité des instruments d'évaluation dans le paradigme constructiviste est un critère négligeable qui, de surcroît, ne pourrait être déterminé étant donné les principes mêmes du paradigme constructiviste. Selon Hein (1991), plus la méthodologie d'administration et de correction d'une évaluation est simple, et moins elle est susceptible de varier en raison de circonstances spécifiques ou du contexte où elle est réalisée, plus il est facile d'établir la fidélité. L'évaluation de toute compétence dans le paradigme constructiviste demande un contexte spécifique et le recours à des connaissances spécifiques. Il ne saurait donc être question de fidélité pour ces types d'évaluation.

Une telle culture à l'endroit des pratiques évaluatives impose cependant de nombreuses contraintes à la validité des instruments et des démarches d'évaluation. Selon Hein (1991), plus les instruments utilisés sont simples et pratiques et plus ils sont conçus pour être administrés à un grand nombre d'élèves simultanément, sur une courte période de temps et en recourant à un minimum de matériel, plus leur validité est douteuse. Le paradigme constructiviste impose donc la reconsidération des critères traditionnels de validité et l'établissement d'une nouvelle hiérarchisation de ces critères.

Le type de validité le plus important est certainement la validité de construit. Selon Pellegrino (1992), il s'agit également du critère le plus controversé et il ajoute que toute pratique évaluative doit être conforme à une théorie sous-jacente. À cet égard, Chittenden (1991) insiste sur le fait que les instruments d'évaluation doivent être élaborés en faisant référence non pas à une norme, mais à une théorie. La validité de construit est définie comme le «degré d'adéquation entre la structure d'un instrument de mesure et le schéma théorique illustrant les caractéristiques comportementales interreliées du trait mesuré» (Legendre, 1985, 634). Le respect de la validité de construit exige la détermination des composantes de chaque compétence et, surtout, de leurs nécessaires interactions pour que la compétence puisse se développer. Dans les phases d'évaluation, le paradigme constructiviste impose le recours à des tâches complètes, complexes et signifiantes. Il demande, de plus, qu'aucune composante ne soit isolée de la compétence qui lui octroie sa signification dans l'expertise cognitive. Par conséquent, toute évaluation de composantes pour elles-mêmes est tout à fait exclue. Elle créerait une certaine forme d'invalidité de construit.

Dans le respect de la validité de construit, il est également nécessaire que les instruments d'évaluation fournissent des informations sur les métamorphoses de chaque compétence de l'élève en rendant explicites ses forces et ses faiblesses cognitives ainsi que métacognitives. Une très grande attention doit être accordée au réel retenu dans les instruments et dans les modalités d'évaluation. Il faut aussi ajouter, bien que cet aspect ait été peu exploré dans le présent texte, que toute démarche évaluative doit fournir des informations sur la perception que l'élève a de ses compé-

tences et de leur contrôlabilité par lui-même. Toujours dans le but de mettre en place des instruments dont la validité de construit est maximale, il est essentiel de prévoir qu'ils doivent fournir des informations cumulatives sur les métamorphoses cognitives de l'élève.

La validité de contenu est une autre valeur métrologique capitale des instruments d'évaluation dans le paradigme constructiviste. Dans le dictionnaire de Legendre (1985, 634), elle est définie comme «le type de validité qui indique le degré de : a) congruence de chacun des items avec l'objectif mesuré; b) représentativité de l'ensemble des items d'un instrument de mesure en regard de ce qu'on veut mesurer.» Pour ce qui est de ce type de validité des instruments d'évaluation dans le paradigme constructiviste, il importe que toutes les questions et les situations soient en conformité avec chaque compétence évaluée. Les questions et les situations sont orientées sur des compétences et non sur des composantes. De là la congruence avec l'objectif mesuré. De plus, au regard de la représentativité de l'ensemble des items, il faut qu'un instrument ou une modalité d'évaluation permette une description exhaustive et adéquate d'une compétence et des interactions de ses composantes. L'évaluation doit également permettre de tirer des conclusions sur la transférabilité de la compétence. Selon cette optique, il serait fort étonnant qu'un seul instrument puisse fournir les informations nécessaires sur le développement d'une compétence donnée.

Une autre forme de validité exigée des instruments d'évaluation dans le paradigme constructiviste est la validité écologique. Jusqu'à maintenant, ce type de validité a plutôt été pris en considération dans le contexte de la recherche. Dans ce secteur d'activités, elle est ainsi définie : «Sorte de validité extrinsèque d'une recherche qui détermine le degré de ressemblance entre les conditions dans lesquelles se trouvent les sujets expérimentaux et celles de la population entière» (Legendre, 1985). Appliquée aux instruments d'évaluation, la validité écologique serait définie comme le degré de ressemblance entre les conditions dans lesquelles se trouvent les élèves dans une situation d'évaluation et celles qui existent à l'extérieur de la classe et de l'institution scolaire. L'école poursuit l'objectif de contribuer au développement d'élèves qui sont capables de «raisonner, de s'adapter et de résoudre des problèmes dans

un environnement en perpétuel changement» (Shepard, 1992). La validité écologique des instruments et des modalités d'évaluation exige que ces habiletés cognitives de haut niveau soient directement prises en considération.

La validité écologique est un critère fondamental des instruments d'évaluation dans une culture de l'évaluation qui se préoccupe des compétences de l'élève et, surtout, de leur transférabilité. Le respect de la validité écologique assure non seulement que les phases d'évaluation aient des retombées significatives pour l'enseignement et l'apprentissage, mais aussi que les conclusions obtenues soient cognitivement et socialement, voire professionnellement, signifiantes.

Une dernière forme de validité que les instruments d'évaluation dans le paradigme constructiviste doivent respecter est la validité déontologique, parfois nommée validité de conséquence. Toutes les évaluations donnent lieu à des conclusions fort importantes pour les personnes évaluées. Les conclusions obtenues ont, entre autres, des retombées directes, fréquemment immédiates, sur le cheminement scolaire de l'élève, sur les choix qui lui sont proposés ainsi que sur les perceptions qui sont développées à son endroit. Dans ce sens, tout acte d'évaluation prend une très grande importance et a beaucoup de signification pour la personne évaluée. C'est selon cette optique que la validité de conséquence ou déontologique est un critère crucial. Cette forme de validité est définie comme le degré d'exactitude avec lequel un instrument ou une modalité d'évaluation détermine les compétences développées. Tout élève a le droit à une évaluation exacte de ses compétences, à une évaluation déontologique de son expertise cognitive et métacognitive. Il faut rappeler ici les conclusions de Gardner (1992) au sujet des tests formels. Selon lui, ces tests conviennent très bien aux personnes qui possèdent des habiletés linguistiques et logiques et qui se sentent à l'aise dans un contexte d'évaluation impersonnel et décontextualisé. Ces tests sont biaisés pour les personnes qui ne possèdent pas ces caractéristiques cognitives et qui ne manifestent leurs compétences que dans des projets continus et des contextes signifiants.

Dans la discussion de la validité de conséquence, certains auteurs (Cronbach, 1988; Linn, Baker et Dunbar, 1991; Messick, 1989) men-

tionnent aussi que les effets des pratiques évaluatives sur les pratiques d'enseignement doivent être considérés très attentivement. Selon ces auteurs, le contenu des instruments d'évaluation proposés aux enseignants influence grandement les compétences et les connaissances sur lesquelles ils décident d'intervenir par la suite. Pour l'enseignant, ce qui est privilégié dans les évaluations correspond à ce qui doit être privilégié dans l'enseignement. La validité de conséquence doit aussi tenir compte des effets des pratiques d'évaluation sur les conceptions et les pratiques de l'enseignant.

CONCLUSION

Le paradigme constructiviste a grandement changé le regard porté sur l'apprentissage et, en conséquence, sur l'enseignement et l'évaluation. L'apprentissage est conçu essentiellement comme la construction graduelle du savoir et cette construction s'opère par le traitement actif des informations à la disposition de l'apprenant. Le savoir construit contient des connaissances déclaratives, des connaissances conditionnelles et des connaissances procédurales organisées et hiérarchisées. Des stratégies cognitives et métacognitives sont aussi intégrées en mémoire à long terme pour la réutilisation et le transfert du savoir. Dans le paradigme constructiviste, l'enseignant poursuit l'objectif avoué de contribuer directement et explicitement au développement de l'expertise cognitive de l'élève. Dans ce cadre, pour reprendre les mots mêmes d'O'Connor (1992), chaque élève est une responsabilité sérieuse.

Afin de participer au développement de l'expertise cognitive de l'élève, l'enseignant met en place des activités complexes, complètes et signifiantes. Le contexte ainsi créé exige l'engagement et la participation active de l'élève. Les évaluations sont conformes à ces pratiques d'enseignement et elles correspondent, selon Gardner (1992), à des problèmes ou à des projets qui intéressent l'élève et qui le motivent à s'engager et à persister. Le paradigme constructiviste, tant pour ce qui est des pratiques d'enseignement que des pratiques d'évaluation, requiert que l'enseignant accorde beaucoup d'attention à l'analyse des tâches proposées à l'élève. Il faut également qu'il puisse analyser le contenu de chaque champ de connaissances dont il a la responsabilité, qu'il soit conscient du chemi-

nement cognitif dans ces champs de connaissances et qu'il puisse prévoir les problèmes que l'élève est susceptible de rencontrer. Dans cette perspective, les exigences de la mise en place des principes du paradigme constructiviste sont nombreuses, variées et, certainement, fort complexes.

RÉFÉRENCES

ANDERSON, J.R. (1983).
The Architecture of cognition. Cambridge (MA) : Harvard University Press.

ANDERSON, J.R. (1985).
Cognitive psychology and its implications (2e édition). New York : Freeman.

AUSUBEL, D.P. (1968).
Educational psychology : A cognitive view. New York : Holt, Rinehart and Winston.

CHITTENDEN, E. (1991).
Authentic assessment, evaluation, and documentation of student performance. *In* V. Perrone (éd.), *Expanding student assessment* (p. 22-32). Alexandria (VA) : Association for Supervision and Curriculum Development.

CRONBACH, L.J. (1988).
Five perspectives on validity argument. *In* H. Wainer et H. I. Braun (éd.), *Test validity* (p. 3-17). Hillsdale (NJ) : Lawrence Erlbaum Associates.

FEUERSTEIN, S. (1969).
The instrumental enrichment method : An outline of theory and technique. Jerusalem : Hadassah-Wizo-Canada Research Institute.

FREDERIKSEN, N. (1990).
Introduction. *In* N. Frederiksen, R. Glaser, A. Lesgold et M.G. Shafto (éd.), *Diagnostic monitoring of skill and knowledge acquisition* (IX-XVII). Hillsdale (NJ) : Lawrence Erlbaum Associates.

GAGNÉ, E.D. (1985).
The cognitive psychology of school learning. Boston : Little, Brown and Company.

GARDNER, H. (1992).
Assessment in context : The alternative to standardized testing. *In* B. R. Gifford et M. C. O'Connor (éd.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (p. 77-121). Boston : Kluwer Academic Publishers.

GIORDAN, A. et DE VECCHI, G. (1987).
Les origines du savoir : des conceptions des apprenants aux concepts scientifiques. Paris : Delachaux et Niestlé.

JONES, B.F., PALINCSAR, A.S., OGLE, D.S. et CARR, E.G. (1987).
Strategic teaching and learning : Cognitive instruction in the content areas. Alexandria (VA) : Association for Supervision and Curriculum Development.

LEGENDRE, R. (1988).
Dictionnaire actuel de l'éducation. Paris : Larousse.

LESGOLD, A., LAJOIE, S., LOGAN, D. et EGGAN, G. (1990).
Applying cognitive task analysis and research methods to assessment. *In* N. Frederiksen, R. Glaser, A. Lesgold et M.G. Shafto (éd.), *Diagnostic monitoring of skill and knowledge acquisition* (p. 325-351). Hillsdale (NJ) : Lawrence Erlbaum Associates.

LIDZ, C.S. (1987).
Dynamic assessment : An interactional approach to evaluating learning potential. New York : Guilford Press.

LINN, R.L., BAKER, E.L. et DUNBAR, S.B. (1991).
Complex, performance-based assessment : Expectations and validation criteria. *Educational Researcher, 20*(8), 15-24.

MARZANO, R.J., BRANDT, R.S., HUGHES, C.S., JONES, B.F., PRESSEISEN, B.Z., RANKIN, S.C. et SUHOR, C. (1988).
Dimensions of thinking : A framework for curriculum and instruction. Alexandria (VA) : Association for Supervision and Curriculum Development.

MESSICK, S. (1989).
Validity. *In* R. L. Linn (éd.), *Educational measurement* (3e édition) (p.13-104). New York : Macmillan

MEYER, B.J.F. (1975).
The organization of prose and its effects on memory. Amsterdam : North-Holland.

MEYER, B.J.F. (1985).
Prose analysis : Purposes, procedures, and problems. *In* B. Britton et J. Black (éd.), *Understanding Expository Text* (p. 11-65). Hillsdale (NJ) : Lawrence Erlbaum Associates.

MEYER, B.J.F. (1987).
Following the author's top-level organization : An important skill for reading comprehension. *In* R. Tierney, P. Andres et J. Mitchell (éd.), *Understanding readers' understanding* (p. 59-77). Hillsdale (NJ) : Lawrence Erlbaum Associates.

MEYER, B.J.F. et RICE, G.E. (1984).
The structure of text. *In* P. D. Pearson (éd.), *Handbook of reading research* (p. 316-342). New York : Longman.

MEYER, B.J.F., BRANDT, D.M. et BLUTH, G.J. (1980).
Use of top-level structure in text : Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16*, 72-103.

MEYER, C.A. (1992).
What's the difference between authentic and performance assessment? *Educational Leadership, 49* (8), 39.

O'CONNOR, M.C. (1992).
Overview : Rethinking aptitude, achievement, and instruction : Cognitive science research and the framing of assessment policy. *In* B. R. Gifford et M. C. O'Connor (éd.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (p. 9-36). Boston : Kluwer Academic Publishers.

O'NEIL, J. (1992).
Putting performance assessment to the test. *Educational Leadership, 49*(8), 14-20.

PELLEGRINO, J.W. (1992).
Understanding what we measure and measuring what we understand. *In* B. R. Gifford et M. C. O'Connor (éd.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (p. 275-301). Boston : Kluwer Academic Publishers.

PERRONE, V. (1991).
Moving toward more powerful assessment. *In* V. Perrone (éd.), *Expanding student assessment* (p. 164-167). Alexandria (VA) : Association for Supervision and Curriculum Development.

RESNICK, L.B. et RESNICK, D.L. (1992).
Assessing the thinking curriculum : New tools for educational reform. *In* B. R. Gifford et M. C. O'Connor (éd.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (p. 37-77). Boston : Kluwer Academic Publishers.

SHEPARD, L.A. (1992).
Commentary : What policy makers who mandate tests should know about the new psychology of intellectual ability and learning. *In* B. R. Gifford et M. C. O'Connor (éd.), *Changing assessments : Alternative views of aptitude, achievement and instruction* (p. 301-329). Boston : Kluwer Academic Publishers.

TARDIF, J. (1992).
Pour un enseignement stratégique : l'apport de la psychologie cognitive. Montréal : Les Éditions Logiques.

WIGGINS, G. (1989).
Teaching to the (authentic) test. *Educational Leadership, 46*(7), 41-50.

WIGGINS, G. (1992).
Creating tests worth taking. *Educational Leadership, 49*(8), 26-34.

ZESSOULES, R. et GARDNER, H. (1991).
Authentic assessment : Beyond the buzzword and into the classroom. *In* V. Perrone (éd.), *Expanding student assessment* (p. 47-72). Alexandria (VA) : Association for Supervision and Curriculum Development.

ÉVALUATION FORMATIVE DES PROCESSUS D'APPRENTISSAGE : LE RÔLE DES RÉGULATIONS MÉTACOGNITIVES

LINDA ALLAL
Université de Genève

INTRODUCTION

Les recherches sur l'évaluation formative des apprentissages se sont développées dans plusieurs directions complémentaires. Après une brève revue centrée sur les développements depuis vingt ans en Europe francophone, un domaine de recherche sera traité de manière plus approfondie. Il s'agit des recherches portant sur l'implication de l'élève dans le processus d'évaluation formative. Une analyse de l'ensemble des régulations qui interviennent dans le système didactique permettra de préciser le rôle de l'autorégulation de l'élève. Dans un cadre conceptuel élaboré à partir des travaux sur la métacognition, les opérations de régulation métacognitives de l'élève seront définies et des propositions porteront sur des modalités d'évaluation formative susceptibles de favoriser la mise en œuvre de ces régulations dans le contexte scolaire. En conclusion, des relations seront établies avec des travaux récents conduits aux États-Unis sur «assessment of cognitive and conative structures» et sur «dynamic assessment».

1. DIRECTIONS DE DÉVELOPPEMENT DES TRAVAUX EUROPÉENS EN ÉVALUATION FORMATIVE

Le concept d'évaluation formative, formulé d'abord par Scriven (1967) dans le domaine de l'évaluation des programmes et des moyens d'enseignement, a été ensuite transposé par Bloom et ses collaborateurs (Bloom, 1968; Bloom, Hastings et Madaus, 1971) au domaine de l'évaluation de l'apprentissage de l'élève. Curieusement, les références explicites à l'évaluation formative sont devenues relativement rares dans les publications américaines concernant *«assessment of learning»*, alors que les travaux centrés sur ce thème ont connu un développement considérable dans les centres de recherche du Canada et de l'Europe francophones.

Je présenterai ici quatre directions de développement qui caractérisent les travaux sur l'évaluation formative menés en Europe depuis vingt ans. Il ne s'agit pas de quatre conceptions alternatives et mutuellement exclusives, mais plutôt de quatre «chantiers de recherche» qui se sont successivement ouverts. Chaque nouveau chantier a apporté un élargissement de perspective, sans que les préoccupations des recherches antérieures soient réellement dépassées.

1.1 Les techniques d'instrumentation

Au départ, l'évaluation formative était abordée surtout sous l'aspect de la confection des outils et des instruments nécessaires à sa mise en application. Cette perspective est reflétée dans le premier article de Bloom (1968) qui propose, dans le cadre d'une pédagogie de maîtrise, une instrumentation constituée de «tests formatifs» suivis de moyens de remédiation. L'accent mis sur les outils et sur les instruments prend son essor avec les multiples exemples de tables de spécifications et de tests à référence critérielle présentés dans le célèbre *Handbook* de Bloom, Hastings et Madaus (1971). Des publications analogues de recueils d'instruments destinés aux enseignants se sont développées également en Europe (Marchandisse et Blampain, 1974, Tourneur, Noël et Honclaire, 1975; SIPRI-ATE (Cahier 3), 1982, entre autres).

Les limites d'une approche de l'évaluation formative centrée trop exclusivement sur l'instrumentation ont été mises en évidence par plusieurs chercheurs. Relevons, en particulier, l'analyse que Cardinet (1986) a faite des problèmes de validité des tests à référence critérielle, et la critique de Bain (1988) de «l'illusion instrumentale» d'une évaluation formative sans ancrage dans les principes de la didactique de la discipline d'enseignement.

1.2 Les fondements théoriques

Parmi les tentatives d'intégrer l'évaluation formative dans un cadre de théorisation plus large quant aux processus d'enseignement et d'apprentissage en jeu dans le contexte scolaire, on peut distinguer trois orientations. Premièrement, la référence aux théories de l'apprentissage permet de concevoir des stratégies d'évaluation formative différenciées;

aussi cherche-t-on à dépasser la stratégie néobéhavioriste de départ proposée par Bloom, en faveur de stratégies fondées sur les théories cognitivistes, et à élargir la conception des processus de régulation – interactive, rétroactive, proactive – relevant de l'évaluation formative (Allal, 1979, 1988). Une deuxième orientation préconise la conceptualisation des démarches de l'évaluation formative à partir des principes théoriques de la didactique de la discipline d'enseignement. Cette approche est particulièrement développée dans les travaux récents en didactique du français (Bain, 1988; Garcia-DeBanc et Mas, 1987; Meyer et Phélut, 1983, entre autres). Selon une troisième orientation théorique basée sur les concepts de la psychologie sociale, l'évaluation formative est considérée avant tout comme un «processus de communication» plus ou moins bien «réussi» dans le contexte des échanges maître-élèves et entre élèves au sein de la classe (Cardinet, 1986, 1988; Weiss, 1991).

1.3 Les pratiques effectives et leur mise en contexte

Les développements dans ce domaine proviennent notamment des recherches qui visent la description et l'analyse des pratiques effectives des enseignants en matière d'évaluation (De Ketele, 1986) : par exemple, le processus implicite de «négociation» qui s'instaure entre le maître et sa classe autour des enjeux de l'évaluation (Chevallard, 1986). En proposant une approche «pragmatique» de l'évaluation formative, Perrenoud (1991) cherche à tirer les meilleurs apports des travaux théoriques existants en vue d'une diversification des «passages à l'acte» dans les classes, sans souci prioritaire de doctrine à respecter. L'analyse des pratiques englobe aussi l'étude des facteurs institutionnels et psychosociologiques qui conditionnent l'attitude de l'enseignant face au problème de l'inégalité, et donc sa manière d'envisager l'évaluation formative (Grisay, 1988; Perrenoud, 1988).

1.4 L'implication de l'élève

Au-delà des procédures d'évaluation formative conçues et gérées par les acteurs adultes du système éducatif, la question se pose de l'implication de l'élève dans les processus évaluatifs qui le concernent.

L'introduction des nouvelles méthodes didactiques a été l'occasion pour plusieurs équipes de recherche, travaillant en étroite collaboration avec des enseignants, de développer des démarches d'autoévaluation et d'évaluation mutuelle entre pairs, ce qui a favorisé la participation active de l'élève au processus de formation (Leselbaum, 1982; Allal, 1984, 1985). Dans la conception de l'évaluation «formatrice» de Nunziati (1990), l'élève s'implique dans chaque étape de construction et de mise en œuvre des outils de référence (critères de réalisation et de réussite, cartes d'étude, dispositifs d'autoévaluation). Récemment, le débat entre chercheurs a porté sur les relations entre les aspects instrumentés et interactifs de l'autoévaluation et sur les apports des recherches relatives à la métacognition (Paquay, Allal et Laveault, 1990).

Cette présentation se situera prioritairement dans le prolongement de la quatrième direction de recherche relative à l'implication de l'élève. Après une analyse de l'ensemble des régulations qui interviennent dans le système didactique de la classe, j'examinerai plus particulièrement les opérations de régulation métacognitive de l'élève, ainsi que les modalités d'évaluation formative qui favorisent la mise en œuvre de ces régulations.

2. LES RÉGULATIONS INTERVENANT DANS LE SYSTÈME DIDACTIQUE

Afin d'élargir le concept étroit de remédiation énoncé par Bloom (1968), on propose depuis plusieurs années de situer l'évaluation formative dans une théorie générale des régulations en situation scolaire (Allal, 1988; Perrenoud, 1991). Je développerai ici un cadre conceptuel (Allal et Saada-Robert, 1992) qui décrit les différents mécanismes de régulation qui permettent au système didactique de fonctionner, c'est-à-dire d'articuler les processus d'enseignement et d'apprentissage (figure 1). Bien que ce schéma soit applicable aux régulations qui interviennent dans l'enseignement de n'importe quelle discipline scolaire, je fournirai des illustrations d'observations faites lors de mes récentes recherches sur les processus de régulation en situation de production écrite à l'école primaire (Allal, Michel et Saada-Robert, en préparation).

2.1 Les régulations orchestrées par l'enseignant

Dans le rectangle de la partie supérieure du schéma figurent les actions de régulation orchestrées par l'enseignant. Le terme «orchestrer» est utilisé pour souligner le fait que l'enseignant agit sur les conditions d'apprentissage en classe; il guide et oriente le travail des élèves, mais il ne peut jamais garantir que son action éducative produise les effets de régulation prévus sur les processus d'apprentissage des élèves. Le terme «orchestrer» est employé aussi en raison de la complexité des «partitions» qui se jouent lorsqu'un maître anime une classe de 20-25 élèves.

À l'intérieur du rectangle, une distinction est faite entre trois sortes de régulations qui proviennent potentiellement de l'action de l'enseignant. Premièrement, lorsque l'enseignant met en place une situation didactique, la structure de la situation (les contenus et les éléments d'habillage de la tâche, les consignes, le matériel, l'organisation interpersonnelle du travail, etc.) peut être source de régulations indirectes, c'est-à-dire de régulations qui ne relèvent pas directement de l'intervention de l'enseignant. Par exemple, dans une séquence didactique qui porte sur la production d'un texte narratif, si le maître demande aux élèves de se mettre «en duo», d'échanger leurs textes et de discuter des améliorations possibles, cette situation peut être incitatrice de régulations (se traduisant en révisions des textes), sans qu'il y ait supervision directe du travail par l'adulte.

Outre les régulations provoquées par la situation didactique, il existe, bien entendu, des régulations qui résultent de l'intervention directe de l'enseignant auprès d'un groupe d'élèves ou d'un individu. À ce propos, il me paraît intéressant d'adopter la distinction faite par Perrenoud (1991) entre deux sortes d'interventions du maître : les interventions relatives à l'activité des élèves et celles qui portent sur les démarches d'apprentissage des élèves.

Une intervention qui vise la gestion de l'activité entreprise en classe est fondée le plus souvent sur des principes pédagogiques d'ordre général, sans formulation d'hypothèses spécifiques concernant la progression individuelle de l'apprentissage. Par exemple, supposons que le «contrat didactique» de la classe (Schubauer-Leoni, 1986) comprend

la règle suivante : en situation de production écrite, l'élève doit utiliser le dictionnaire pour vérifier l'orthographe d'un mot douteux avant de demander l'aide de l'enseignant. Dans ce cas, si l'élève demande au maître : «Monsieur, est-ce juste?», et que celui-ci répond : «As-tu regardé dans le dictionnaire?», l'intervention de l'enseignant vise une régulation au niveau de l'activité de l'élève; l'intervention est faite sans diagnostic (le maître n'a pas lu le texte de l'élève), et aurait été essentiellement la même quel que soit l'élève qui ait posé la question.

Malgré le caractère «rudimentaire» de cette forme de régulation, peu ou pas différenciée selon la particularité du cheminement de l'élève, son rôle dans la gestion de l'enseignement peut être néanmoins positif, et ceci pour deux raisons. Premièrement, si le maître ne parvient pas à développer un certain nombre de «routines» qui permettent la régulation globale de l'activité des élèves de sa classe, il ne sera jamais en mesure de consacrer une partie de son temps aux interventions plus personnalisées auprès de tel ou tel élève. Deuxièmement, des régulations au niveau de l'activité de l'élève ont l'avantage de limiter un interventionnisme parfois trop envahissant, et de favoriser ainsi une certaine autonomie chez les élèves qui apprennent peu à peu à intégrer les «règles de vie de classe» à leur fonctionnement, plutôt que de dépendre de la guidance de proche en proche de l'enseignant.

L'intervention directe et individualisée auprès d'un élève, ou d'un petit groupe, reste toutefois une source puissante de régulations en situation scolaire [voir, entre autres, les études de Bloom (1984) sur l'impact du préceptorat]. Dans ce cas, l'action de l'enseignant s'appuie sur une hypothèse interprétative concernant la progression de l'élève dans un cursus d'apprentissage. Le but de l'intervention est de favoriser – par des questions, par des suggestions, par des contre-exemples – la réflexion de l'élève et de stimuler sa recherche active de solutions aux difficultés rencontrées. Afin d'illustrer la régulation qui résulte de ce type d'intervention, on peut prendre le cas d'un mini-entretien réalisé par l'enseignant avec un élève qui vient de terminer le brouillon d'un texte : partant de l'affirmation de l'élève que «la fin ne marche pas», le maître posera des questions pour aider l'élève à préciser l'effet qu'il voudrait produire, les mots (éléments lexicaux, organisateurs) qui marqueraient mieux la conclusion de l'histoire, etc.

Les flèches dans la Figure 1 entre le rectangle (régulations orchestrées par l'enseignant) et l'ellipse (régulations intégrées aux conduites des élèves) sont bidirectionnelles. Cela traduit le fait que même si c'est l'enseignant qui met en place la situation didactique et initie en grande partie les interactions maître-élèves, les conduites des élèves conditionnent fortement l'action du maître. Même une leçon traditionnelle ne fonctionnerait pas si les élèves ne jouaient pas le jeu du scénario «questions/réponses». Cette codétermination des conduites du maître et des élèves est encore plus forte dans les séquences typiques des nouvelles didactiques qui exigent une participation accrue des élèves (Bain et Schneuwly, 1987).

2.2 Les régulations gérées par les élèves

Dans la partie inférieure de la Figure 1, à l'intérieur de l'ellipse, les flèches entre élèves symbolisent les régulations qui résulte de leurs interactions. Ces interactions englobent non seulement les échanges en face-à-face entre élèves, mais aussi les effets sur les conduites des uns et des autres dus à la présence de leurs camarades. Ainsi, un élève peut être amené à allonger un texte, soit parce qu'un camarade lui a dit que «C'est trop court; on ne comprend pas la fin.», soit parce qu'il constate, au moment où il pensait avoir terminé, que toute la classe continue à écrire et qu'il reste encore dix minutes jusqu'à la récréation. Ainsi, les effets de régulation de l'action du maître, telle une consigne relative à la longueur du texte, sont toujours médiatisés par le réseau des interactions entre les élèves.

En outre, les interactions entre élèves ont leurs effets propres. Des recherches effectuées sur des situations de jeu dans l'enseignement de la mathématique ont démontré qu'un processus de régulation interactive, impliquant des contrôles et des suggestions réciproques entre élèves, peut s'instaurer dès les premiers degrés primaires, surtout si la structure de la situation comporte des éléments incitateurs (Allal, 1985). Dans l'enseignement du français, des situations de régulation interactive entre pairs peuvent se justifier encore plus aisément, étant donné qu'elles rejoignent les objectifs de communication de la discipline.

Les modalités décrites jusqu'ici constituent en fait des potentialités de régulation, c'est-à-dire des aspects des situations, des interventions du maître, des interactions entre élèves, qui sont susceptibles de favoriser un processus d'autorégulation chez chaque élève (symbolisé par la flèche en boucle du «retour sur soi»). Autrement dit, les modalités situationnelles et interpersonnelles n'auront d'impact sur l'apprentissage de l'élève que dans la mesure où elles s'intègrent aux régulations propres à l'individu. L'autorégulation, en tant que dimension fondamentale des processus cognitifs, est forcément présente à tout moment en chaque apprenant. Même lorsque l'élève n'écoute plus la leçon ou la discussion entre des camarades, ses mécanismes d'autorégulation fonctionnent et lui permettent, par exemple, de faire semblant d'écouter. Ce qui nous intéresse dans le contexte scolaire n'est donc pas de savoir si l'élève régule ses démarches d'apprentissage (il le fait de toute façon, d'une manière ou d'une autre), mais plutôt de déterminer comment promouvoir des situations didactiques, des interventions du maître, des interactions entre pairs, qui favorisent une autorégulation génératrice d'apprentissage chez chaque élève. Dès lors, il s'agit d'étudier les régulations métacognitives de l'individu en les situant dans une dynamique sociocognitive et dans un contexte didactique définis.

Dans cette présentation de la Figure 1, on remarquera que les processus de régulation sont symbolisés dans les flèches du schéma, mais que le concept d'évaluation formative n'y figure pas explicitement. Il faut le situer, me semble-t-il, dans les conditions qui entourent les flèches, c'est-à-dire dans les procédés d'observation, d'interprétation et de prise de décision inhérents à toute évaluation (ou autoévaluation) et qui orientent et stimulent les régulations. Étant donné que l'effet d'un procédé d'évaluation (d'un test, d'un entretien maître-élève, d'une fiche d'appréciation) dépend inévitablement de son intégration aux processus d'autorégulation de l'élève, une théorisation plus approfondie de ces processus est nécessaire afin de poser des points de repère pour la création de stratégies optimales d'évaluation formative. C'est dans ce but que j'examinerai maintenant les apports des recherches sur la métacognition.

3. RÉGULATIONS MÉTACOGNITIVES ET MODALITÉS D'ÉVALUATION FORMATIVE FAVORISANT LEUR MISE EN ŒUVRE

Dans la perspective qui a été définie par Flavell (1976, 1979) et par Brown (1978, 1987), la métacognition comprend deux dimensions principales : a) les métaconnaissances (connaissances que le sujet possède de ses propres processus de pensée et de ceux d'autrui), et b) les régulations métacognitives ou les processus d'autorégulation que le sujet exerce, plus ou moins explicitement, sur son activité cognitive. En analysant cette deuxième dimension, je m'appuierai sur un cadre conceptuel développé ailleurs (voir Allal et Saada-Robert, 1992, pour une revue systématique des recherches sur lesquelles notre approche théorique est fondée).

3.1 Les opérations de régulation métacognitive

Quand l'apprenant est confronté à une tâche à réaliser, son fonctionnement cognitif implique l'activation d'un «espace de travail» qui comprend trois composantes : le réseau fonctionnel des représentations qui fournit un cadre d'organisation pour l'ensemble de son activité face à la tâche; les processus de production qui assurent l'élaboration d'une séquence de procédures qui se traduisent en conduites de réalisation de la tâche; les opérations de régulation métacognitive qui constituent une sorte d'interface entre représentations et processus de production.

Trois opérations de régulation métacognitive sont définies de la manière suivante.

1) *L'anticipation.* — Cette opération traduit l'organisation des représentations du sujet en orientations de l'action, et assure ainsi le guidage de proche en proche des processus de production.

2) *Le contrôle.* — Il s'agit d'une opération de *«monitoring»* au sens donné à ce terme dans les théories de traitement de l'information; cette opération implique un processus continu de comparaison entre un état donné (l'état présent constaté par le mécanisme de contrôle) et un état-but à atteindre (référentiel évoqué par l'opération d'anticipation); dans le cas d'une tâche complexe, le contrôle peut porter au préalable sur la compatibilité de plusieurs orientations d'action entre elles, telles qu'elles sont évoquées par l'anticipation.

3) *L'ajustement.* — Cette opération est la conséquence de l'opération de contrôle; si le monitoring met en évidence une divergence entre l'état présent et l'état-but (feedback négatif), l'opération d'ajustement introduit une modification ou une réorientation des processus de production; si, au contraire, le contrôle montre la convergence de l'état présent et l'état-but (feedback positif), les processus de production se poursuivent dans la même orientation (ajustement nul), ou cessent si le but final est atteint; l'ajustement peut viser en outre à renforcer la cohérence fonctionnelle du réseau des représentations.

Les trois opérations de régulation ont un ordre séquentiel : l'anticipation, activée par les représentations, sert de référence au contrôle qui, lui, provoque l'ajustement; ce dernier incite en retour une nouvelle anticipation, et ainsi de suite. Mais ce fonctionnement peut se traduire en récurrence provisoire de boucles réduites, soit en une sous-boucle d'anticipations-contrôles répétées lorsque que le monitoring n'appelle aucun ajustement (feedback positif), soit en une sous-boucle de contrôles-ajustements répétés guidés par une même anticipation.

Lors de la réalisation de n'importe quelle tâche, les trois opérations de régulation fonctionnent toujours de manière implicite, en articulant les représentations et les processus de production (régulations dites *on line*). Mais ces opérations peuvent se dérouler aussi de manière explicite et consciente. L'explicitation peut émerger aussi bien en cours de production — par exemple, lorsque l'élève se donne une consigne mentale du type «attention à ne pas répéter encore une fois ce mot» — qu'en dehors du moment de production (*off line*) — par exemple, lorsque l'élève relit un texte terminé dans le but de relever et de modifier d'éventuelles répétitions lexicales.

3.2 Mise en œuvre des régulations et des modalités d'évaluation formative

Ayant défini les opérations qui constituent la régulation métacognitive, je traiterai maintenant de leur mise en œuvre en situation scolaire. Je distinguerai trois niveaux de mise en œuvre en esquissant chaque fois des modalités d'évaluation formative susceptibles de soutenir et de promouvoir des processus de régulation qui favorisent l'apprentissage.

Niveau 1 — Régulations «on line» intégrées au processus de réalisation de la tâche. Il s'agit de régulations qui agissent sur le plan de la pensée représentationnelle, mais qui impliquent presque toujours, dans le contexte scolaire, l'interaction de l'élève avec un support extérieur (matériel didactique, documents de référence, etc.). À ce niveau, l'évaluation formative se traduit surtout par l'implication des élèves dans le développement d'outils d'autoévaluation, et par leur appropriation progressive des outils. Il s'agit de favoriser les processus d'intériorisation et de routinisation des démarches de régulation soutenues initialement par les outils. Des recherches récentes montrent qu'une telle approche est possible notamment à l'aide d'environnements informatisés qui stimulent les stratégies métacognitives des élèves (Salomon, Globerson et Guterman, 1989).

Niveau 2 — Régulations en relation avec la gestion de la tâche. Les régulations à ce niveau portent sur la gestion d'un «produit provisoire» en fonction des exigences et des conditions de réalisation matérielles et sociales de la tâche : par exemple, relire un travail pour vérifier non seulement sa correction mais aussi son adéquation vis-à-vis du destinataire; demander à l'enseignant des ouvrages de référence afin de compléter l'information présentée dans un rapport de recherche; discuter avec des coéquipiers des meilleures représentations graphiques à retenir pour un poster préparé en commun. Les apports de l'évaluation formative au niveau de la gestion de la tâche peuvent se situer sur deux plans :

1) celui des *outils* qui favorisent un recul critique vis-à-vis de sa propre production, ou qui stimulent une meilleure exploitation des ressources à disposition (fiches de contrôle ou de vérification, fiches-guide qui suggèrent des ouvrages de référence, ou des pistes à suivre en cas de difficulté);

2) celui de la *structuration de la situation* didactique qui prévoit des occasions multiples de confrontation d'opinions ou de démarches de travail différentes, aussi bien dans une perspective d'évaluation mutuelle (entre pairs) que dans celle d'une coévaluation (autoévaluation de l'élève confrontée à l'évaluation de l'enseignant).

Des exemples de modalités d'évaluation formative, aussi bien instrumentées que non instrumentées, sont fournis dans Allal (1988) et dans Leselbaum (1982).

Niveau 3 — Régulations liées à la gestion des relations entre tâches et contextes didactiques plus généraux. À ce troisième niveau, il s'agit de régulations qui dépassent le cadre strict de la réalisation d'une tâche et qui concernent les relations entre les multiples tâches – relevant souvent de disciplines diverses – que l'élève doit gérer en tenant compte des règles générales de la vie de classe. Ce troisième niveau de régulation est présent surtout dans les classes caractérisées par un mode d'organisation relativement dynamique où tout n'est pas dirigé de moment en moment par l'enseignant mais où, au contraire, les élèves doivent prendre des décisions concernant la gestion du temps, comme ordonner les tâches à effectuer dans une période de «travail libre», et décider de leur durée, concernant la dimension interpersonnelle du travail, comme demander de l'aide ou non à un camarade, concernant l'exploitation des ressources à disposition, comme savoir comment utiliser la bibliothèque de l'école. Parmi les modalités d'évaluation formative favorables aux régulations métacognitives de niveau 3, on peut citer notamment la mise en place de systèmes de «contrat pédagogique» ou de «plan de semaine» (cf. Jarnieu et Wyler, 1981). Ces systèmes offrent un cadre dans lequel l'élève peut apprendre à gérer la complexité des relations entre tâches et contextes : planifier une séquence d'activités, contrôler sa progression, réviser ses projets de départ.

CONCLUSION

En conclusion, je souhaiterais évoquer deux perspectives présentées dans des publications récentes aux États-Unis; ces perspectives susceptibles d'enrichir notre compréhension des régulations métacognitives et de suggérer de nouvelles conceptions de l'évaluation formative des processus d'apprentissage.

Dans la perspective de *«cognitive and conative assessment»* que Snow (1989, 1990) a élaborée, il est intéressant de relever la place accordée à l'autorégulation. Selon ce modèle, les régulations métacognitives se situent à la frontière entre les compétences essentiellement cognitives

de l'élève (*conceptual structures, procedural skills, learning strategies*) et ses dispositions motivationnelles. Dans l'optique d'un élargissement des approches d'évaluation au-delà des tests centrés sur la mesure des connaissances, Snow aborde l'autorégulation surtout en tant qu'objet d'évaluation. Cependant, les nouveaux types de procédures qu'il préconise, dont la technique du «*teach-back*» (demander à l'élève d'enseigner un concept à une autre personne), accordent forcément un rôle à l'autorégulation dans le fonctionnement du processus d'évaluation.

Une deuxième perspective pertinente est celle de l'appréciation dynamique («*dynamic assessment*») développée par Campione et Brown (1990) sur la base de la théorie de Vygotsky (1978) de la médiation sociale de l'apprentissage. Résumée de manière très succincte, l'idée centrale est de faire porter l'évaluation non seulement sur l'état des connaissances et des «*skills*» acquis, mais aussi sur les compétences *émergentes* de l'élève, c'est-à-dire sur les compétences situées dans ce que Vygotsky a appelé la «zone proximale de développement» de l'apprenant. Ainsi, en conduisant une appréciation dynamique, on relève non seulement les items (ou tâches) que l'élève est capable de réussir, comme dans un test traditionnel, mais aussi les items (ou tâches) que l'élève parvient à réaliser si l'expérimentateur (ou l'enseignant) lui fournit de l'aide. L'aide est conçue, dans cette approche, sous forme de «*hints*», c'est-à-dire de questions et de suggestions susceptibles de «relancer» la réflexion de l'élève et de favoriser des démarches actives d'exploration et de structuration de sa part. Une perspective similaire qui vise une évaluation «fonctionnelle et dynamique» a été formulée par des chercheurs québécois (Bourassa et Leblanc, 1991).

L'appréciation dynamique offre la possibilité, me semble-t-il, de mieux saisir certains aspects des régulations métacognitives. Dans une tâche maîtrisée, les processus de régulation sont devenus automatisés et sont, par conséquent, difficilement accessibles à l'observateur. En revanche, lorsqu'on propose à l'élève une tâche qui relève de sa zone proximale de développement, des manifestations explicites de ses processus de régulation ont tendance à paraître. En observant les réactions de l'élève aux «*hints*», on relève des conduites révélatrices de ses opérations de régulation métacognitive. Comment interprète-t-il, transforme-t-il le «*hint*»

proposé? Comment s'en sert-il pour planifier, pour contrôler ainsi que pour ajuster ses démarches face à la tâche?

Les perspectives esquissées ici offrent de nouvelles pistes pour le développement des recherches sur l'évaluation formative. En ce qui concerne l'autorégulation de l'élève, il sera particulièrement important de mettre en œuvre une double direction d'investigation portant, d'une part, sur l'*autorégulation en tant qu'objet d'une évaluation formative* et d'autre part, sur l'*autorégulation en tant que processus impliquée dans la conduite d'une évaluation formative* des diverses compétences de l'apprenant.

RÉFÉRENCES

ALLAL, L. (1979).
Stratégies d'évaluation formative : conceptions psychopédagogiques et modalités d'application. *In* L. Allal, J. Cardinet et P. Perrenoud (éd.), *L'évaluation formative dans un enseignement différencié*. Berne : Lang.

ALLAL, L. (1984).
Vers une implication active de l'élève dans le processus d'évaluation formative. *Éducateur, 3*, 22-26.

ALLAL, L. (1985).
Processus de régulation interactive dans le cadre d'un jeu de mathématique à l'école primaire. *Revue internationale de psychologie appliquée, 34*, 83-104.

ALLAL, L. (1988).
Vers un élargissement de la pédagogie de la maîtrise : processus de régulation interactive, rétroactive et proactive. *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise*. Neuchâtel : Delachaux et Niestlé.

ALLAL, L. MICHEL, Y. et SAADA-ROBERT, M. (en préparation).
Processus de régulation chez l'élève en situation de production d'un texte informatif.

ALLAL, L. et SAADA-ROBERT, M. (1992).
La métacognition : cadre conceptuel pour l'étude des régulations en situation scolaire. *Archives de Psychologie, 60*, 265-296.

BAIN, D. (1988).
Pour une formation à l'évaluation formative intégrée à la didactique. *In* M. Gather Thurler et P. Perrenoud (éd.), *Savoir évaluer pour mieux enseigner : Quelle formation des maîtres?* (Cahier 26). Genève : Service de la recherche sociologique.

BAIN, D. et SCHNEUWLY, D. (1987).
Vers une pédagogie du texte. *Français Aujourd'hui, 79*, 13-24.

BLOOM, B.S. (1968).
Learning for mastery. *Évaluation Comment, 1*(2), 1-12.

BLOOM, B.S. (1984).
The two-sigma problem : The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

BLOOM, B.S. HASTINGS, J.T. et MADAUS, G.F. (1971).
Handbook on formative and summative evaluation of student learning. New York : Mc Graw-Hill.

BOURASSA, M. et LEBLANC, R. (1991).
Révision de l'évaluation : une démarche fonctionnelle et dynamique, *Mesure et évaluation en éducation, 13*(4), 23-47.

BROWN, A.L. (1978).
Knowing when, where and how to remember : A problem of metacognition. *In* R. Glaser (éd.), *Advances in instructional psychology*. Hillsdale : Erlbaum.

BROWN, A.L. (1987).
Metacognition, executive control, self-regulation and other more mysterious mechanisms. *In* F.E. Weinert et R.H. Kluwe (éd.), *Metacognition, motivation and understanding*. Hillsdale : Erlbaum.

CAMPIONE, J.C. et BROWN, A.L. (1990).
Guided learning and transfer : Implications for approaches to assessment. *In* N. Fredericksen, R. Glaser, A. Lesgold et M.G. Shafto (éd.), *Diagnostic monitoring of skill and knowledge acquisition*. Hillsdale : Erlbaum.

CARDINET, J. (1986).
Évaluation scolaire et pratique. Bruxelles : De Bœck.

CARDINET, J. 1988).
La maîtrise : communication réussie. *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise*. Neuchâtel : Delachaux et Niestlé.

CHEVALLARD, Y. (1986).
Vers une analyse didactique des faits d'évaluation. *In* J.-M. De Ketele (éd.), *L'évaluation : approche descriptive ou prescriptive?* Bruxelles : De Bœck.

DE KETELE, J.-M. (éd.). (1986).
L'évaluation : approche descriptive ou prescriptive? Bruxelles : De Bœck.

FLAVELL, J.H. (1976).
Metacognitive aspects of problem solving. *In* L.B. Resnick (éd.), *The nature of intelligence*. Hillsdale : Erlbaum.

FLAVELL, J.H. (1979).
Metacognition and cognitive monitoring : A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906-911.

GARCIA-DEBANC, C et MAS, M. (1987).
Évaluation des productions écrites. *In* J.-L. Chiss, J.-P. Laurent, J.-C. Meyer, H. Romian, et B. Schneuwly (éd.), *Apprendre/enseigner à produire des textes écrits*. Bruxelles : De Bœck.

GRISAY, A. (1988).
La pédagogie de maîtrise face aux rationalités inégalitaires des systèmes d'enseignement. *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise*. Neuchâtel : Delachaux et Niestlé.

JARNIEU, O. et WYLER, M. (1981).
Le plan de travail. Textes libres. Genève : Projet RAPSODIE — École Lignon, N° 5.

LESELBAUM, N. (1982).
Autonomie et auto-évaluation. Paris : Economica.

MARCHANDISSE, G. et BLAMPAIN, D. (1974).
Techniques d'évaluation formative en langue maternelle. Bruxelles : Ministère de l'Éducation nationale de la Culture française.

MEYER, J.C. et PHÉLUT, J.L. (1983).
Apprendre à écrire le français au collège. Lyon : Chronique sociale.

NUNZIATI, G. (1990).
Pour construire un dispositif d'évaluation formatrice. *Cahiers pédagogiques, 280*, 47-64.

PAQUAY, L., ALLAL, L. et LAVEAULT, D. (1990).
L'auto-évaluation en question(s) : propos pour un débat. *Mesure et évaluation en éducation, 13*(3), 5-26.

PERRENOUD, P. (1988).
La pédagogie de maîtrise : une utopie rationaliste? *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise*. Neuchâtel : Delachaux et Niestlé.

PERRENOUD, P. (1991).
Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation, 13*(4), 49-81.

SALOMON, G., GLOBERSON, T. et GUTERMAN, E. (1989).
The computer as a zone of proximal development : Internalizing reading-related metacognitions from a reading partner. *Journal of Educational Psychology, 81*, 620-627.

SCHUBAUER-LEONI, M.-L. (1986).
Le contrat didactique : un cadre interprétatif pour comprendre les savoirs manifestés par les élèves en mathématique. *European Journal of Psychology of Education, 1*, 139-153.

SCRIVEN, M. (1967).
The methodology of evaluation. *AERA Monograph Series on Evaluation, 1*, 39-83.

SIPRI-ATE. (1982).
Appréciation du travail de l'élève. Cahier *3 : Catalogue des instruments* (J. Weiss, éd.). Neuchâtel : Institut Romand de recherches et de documentation pédagogiques.

SNOW, R.E. (1989).
Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*(9), 8-14.

SNOW, R.E. (1990).
New approaches to cognitive and conative assessment in education. *International Journal of Educational Research, 14*, 455- 473.

TOURNEUR, Y. NOËL, Y. et HONCLAIRE, B. (1975).
Liste des objectifs, épreuves d'évaluation et outils de rattrapage en mathématiques. Bruxelles : Ministère de l'Éducation nationale et de la Culture française.

VYGOTSKY, L.S. (1978).
Mind in society : The development of higher psychological process. Cambridge (MA) : Harvard University Press.

WEISS, J. (éd.). (1991).
L'évaluation : problème de communication. Cousset, Fribourg : DelVal.

RÉGULATION DES APPRENTISSAGES, CONTEXTE DE MESURE ET ÉVALUATION INFORMATISÉE

CLÉMENT DASSA
Université de Montréal

INTRODUCTION

Le champ d'étude de la mesure et de l'évaluation en éducation s'est, depuis quelques années, érigé en discipline autonome qui porte sur un grand nombre d'objets : les institutions, les programmes, l'enseignement et les apprentissages. Historiquement colonisée par la psychologie et plus précisément par la psychométrie, la mesure et l'évaluation des apprentissages scolaires a grandement évoluée au cours des trente dernières années.

Aucune conception de la mesure en éducation qui prétend la fonder sur une théorie, c'est-à-dire sur une conception qui décrit et prédit des phénomènes, qui se prête à une épreuve d'hypothèse à partir de données empiriques et qui génère des informations non triviales, ne résiste à un examen sérieux. En fait, plutôt que d'être fondée sur une théorie, la mesure se réfère soit à des modèles statistiques qui permettent de résumer les réponses aux items, soit à une méthodologie statistique qui visent à résumer des scores d'items (Goldstein, 1991). Or l'interprétation de la mesure, préalable à toute évaluation, nécessite une mise en contexte qui reflète non seulement les buts et les objectifs, mais également les contraintes de divers ordres ainsi que les pratiques éducatives. Idéalement la mesure elle-même devrait être ouverte aux mêmes contextes qui président à une évaluation rigoureuse et juste des apprentissages.

Quant à l'évaluation, la tendance actuelle est à la transparence. L'évaluation étale de plus en plus ses prémisses, c'est-à-dire les valeurs qui sous-tendent les buts des apprentissages scolaires. Au niveau du discours officiel, les principes et les méthodes d'évaluation se fondent de plus en plus sur les valeurs de justice et d'équité. Ces valeurs sont souvent implicites lorsqu'elles ne sont pas clairement formulées dans

les énoncés de politique et les plans d'action officiels[1]. Sous l'impulsion de Tyler qui a élargi le champ et les buts de l'évaluation en milieu scolaire ainsi que sous l'influence de Bloom et de l'École de Chicago qui ont élaboré un paradigme d'évaluation des apprentissages centré sur le concept d'objectifs de comportement, ces valeurs ont été à l'origine d'un grand effort de communication et de formalisation. En effet, «plus ce qui est prévu est exprimé clairement, et plus ce qui est réalisé est analysé formellement, mieux on sera éclairé sur la réalité et sur la qualité de l'apprentissage[2]». Au Québec, on peut affirmer que le modèle dominant d'évaluation des apprentissages, fondé sur une plus grande recherche de rigueur et de justice, s'inscrit dans le cadre d'une évolution marquée au sceau des avancées théoriques de Tyler et de Bloom.

1. LE CONTEXTE ACTUEL DE L'ÉVALUATION

Sans vouloir tracer une analyse historique de l'évolution des concepts, des principes et des pratiques de l'évaluation des apprentissages – cette analyse reste à faire – il nous paraît évident que la préoccupation actuelle du développement intégral de l'individu alliée à une recherche de rationalisation des ressources entraînent des formes d'évaluation propres à guider les apprentissages des élèves ainsi que les stratégies d'enseignement en cours de réalisation. Qui plus est, l'emphase mise aujourd'hui sur les prédicteurs de la réussite dans la vie nous renvoie à une recherche de l'optimisation de la réussite scolaire. Lorsque cette préoccupation de réussite est axée sur le cheminement personnel, elle mène à des formes d'évaluation formative et critériée centrées sur l'identification des difficultés individuelles et leur remédiation (Conseil supérieur de l'éducation, 1987). D'une certaine façon, cette recherche de la réussite s'inscrit dans la prolongation des grands principes de démocratisation de l'enseignement : les systèmes d'éducation qui prônaient l'égalité des chances d'accès visent, à présent, à optimiser les chances de réussite. Évolution profonde et inévitable de nos sociétés ou bien préoccupation

1 Ceci est notamment le cas au Québec où la «Politique générale d'évaluation pédagogique» des secteurs préscolaire, primaire et secondaire, en vigueur depuis le début des années quatre-vingt, pose la justice et l'égalité comme valeurs sous-jacentes à l'évaluation pédagogique.

2 Politique générale d'évaluation pédagogique, ministère de l'Éducation, Direction générale du développement pédagogique. Québec : Document 16-7500, p. 6.

ponctuelle du rendement de l'éducation stimulée par la valorisation renouvelée de la compétitivité économique à l'échelle mondiale? Qu'ils soient indépendants ou en interaction, ces deux facteurs semblent préoccuper de plus en plus un grand nombre d'intervenants de la classe politique ainsi que des milieux des affaires, universitaires et éducatifs. Les nombreuses études et forums[3] en témoignent. Prise entre la course à la productivité collective et la nécessité de préserver, voire de promouvoir, les droits individuels de l'élève ou de l'étudiant à poursuivre un itinéraire personnel garant de son développement général (Ministère de l'Éducation du Québec, 1979), l'école cherche les moyens d'optimiser le rendement scolaire tout en visant à favoriser le «développement holistique, intégral ou global» (Conseil supérieur de l'éducation, 1990).

Dans ce contexte, si l'évaluation se veut *formative* dans le but de guider l'élève, l'enseignement, lui, se doit d'être différencié, c'est-à-dire d'être «rendu différent selon les élèves par une adaptation qui se situe essentiellement au niveau des méthodes, des cheminements, des exemples, etc.» (Cardinet, 1981). Idéalement, cette adaptation constante devrait prendre la forme d'un dialogue permanent entre l'enseignant et l'élève, d'un «acte global de communication interpersonnelle» où l'interaction entre ces deux acteurs servirait de moteur à la régulation des apprentissages. Certes, cette communication ne peut se réaliser indépendamment des divers ordres de contraintes; en fait, elle se trouve fortement structurée par ces derniers qui, de plus en plus, sont façonnés par une recherche de l'efficacité.

2. LA PLACE DE L'ÉVALUATION FORMATIVE DANS UNE PÉDAGOGIE EFFICACE

Le milieu scolaire n'échappe pas aux nombreux jugements d'efficacité, comme l'indiquent les travaux de Brunet, Brassard et Corriveau (1991). Bien qu'abondent les écrits concernant l'efficacité des organisations abondent, on ne peut en dégager de consensus, ni sur la définition du concept, ni sur la façon de l'évaluer. Cette situation peut s'expliquer, d'une part, par l'état des connaissances sur les fondements mêmes des organisations sociales (Kanter et Brinkerhoff, 1981) et, d'autre part,

3 À ce sujet, voir les travaux et documents produits par Forum entreprises-universités, dont «Nous accomplir : Apprendre pour l'avenir», Montréal, 1991, 30 p.

par le caractère pluridisciplinaire de ce domaine de connaissance (Morin, 1989). Ainsi, lorsqu'on étudie le couple apprentissage-enseignement, l'enseignement efficace consiste à optimiser, par le biais d'interventions pédagogiques, la qualité des apprentissages relatifs aux objectifs pédagogiques visés. Or, l'environnement naturel où se réalise l'essentiel de ces interventions, la classe, est organiquement lié à l'école qui, elle, est directement tributaire de la société. Dans ce contexte, on peut définir l'efficacité comme la capacité d'une organisation d'engendrer des actions et des résultats acceptables pour les personnes et les groupes d'intérêts qui interagissent avec elle (Barnard, 1938; Cummings, 1977; Pfeffer et Salancik, 1978). Il convient donc d'aborder l'étude de l'efficacité selon une problématique axée sur l'interdisciplinarité et circonscrite à l'école en tant que point focal de l'intégration du pédagogique à l'organisationnel et à l'écologique. Du fait que les critères qui représentent l'efficacité pédagogique selon les personnes ou les groupes, et selon l'usage qu'ils en font, l'efficacité est une norme de la performance organisationnelle, inévitablement liée aux valeurs des personnes et des groupes qui l'utilisent (Keeley, 1984; Van de Ven et Ferry, 1980; Zammuto, 1984). Par voie de conséquence, cette notion ne peut avoir de signification en soi; elle ne peut en avoir que pour ceux qui s'y réfèrent dans leurs rapports avec les organisations; d'où une multitude de représentations de cette notion (Cameron et Whetten, 1983; Morin, 1989).

C'est dans ce cadre que se pose le problème de la recherche d'une pédagogie efficace ainsi que de celui de la place de l'évaluation formative dans celle-ci. Depuis une vingtaine d'années, les contributions théoriques et pratiques dans le domaine de l'évaluation formative se sont accumulées à un rythme accéléré. Mais qu'en est-il aujourd'hui de nos connaissances des conditions et des facteurs qui permettent de l'intégrer à l'enseignement réel et à l'apprentissage vécu en classe? Par enseignement réel, nous entendons les prestations pédagogiques qui portent sur un contenu, des objectifs et des interventions qui expriment les visées des programmes officiels mais aussi et surtout le quotidien dans la salle de classe tel qu'il est façonné par les multiples contraintes pédagogiques, organisationnelles et institutionnelles. Entre la réalité et le projet, il convient de trouver des pistes d'intégration de l'évaluation formative à l'écologie de la classe.

3. L'ÉLARGISSEMENT DE LA RÉGULATION

Une certaine approche visant cette intégration s'est longtemps imposée aux États-Unis. Elle repose sur la notion de boucle de régulation des apprentissages et consiste en un test critérié administré à la suite d'une période d'enseignement, test suivi d'interventions de nature corrective. Comme le montrent les études récentes (Kulik, C.L., Kulik, J.A. et Bangert-Drowns, 1990; Slavin, 1987), cette version instrumentée[4] de la pédagogie de maîtrise est performante à tous les niveaux d'étude : on décèle des effets positifs significatifs (en moyenne, la grandeur de l'effet est de l'ordre de 0,5 à 0,6) en ce qui concerne aussi bien les variables cognitives (relevant du rendement scolaire) qu'affectives (portant essentiellement sur les attitudes à l'égard des contenus visés et de l'enseignement de ces contenus). L'étude de Kulik, C.L., Kulik, J.A. et Bangert-Drowns (1990) souligne le fait que cette approche exige de la part de l'apprenant plus de temps consacré à l'étude, alors que Slavin (1990) attribue l'augmentation de la performance des étudiants au fait que l'approche incite les enseignants à centrer leurs prestations sur les objectifs visés. Plus modéré dans son appréciation de cette approche, ce dernier auteur rejette toujours l'hypothèse que la pédagogie de maîtrise accélère le rendement aux ordres primaire et secondaire.

Dans de tels systèmes axés sur l'efficacité, il convient donc de suivre une approche pragmatique où l'évaluation formative n'importe que dans la mesure où elle permet de maximiser les apprentissages scolaires en les rendant plus faciles, plus rapides et plus performants. Il importe également, ainsi que le souligne Perrenoud (1991), de remonter des effets attendus sur le processus d'apprentissage «à l'intervention régulatrice qui les produits et, en amont encore, aux observations et aux représentations qui guident cette intervention». La mise en contexte de la mesure devient alors une condition essentielle à toute évaluation des apprentissages surtout de nature formative.

---

4 L'approche instrumentée se subdivise en deux types de procédures. Dans la procédure LFM («Learning for Mastery»), inspirée de Bloom, les étudiants progressent tous au même rythme contrôlé par l'enseignant. Par contre, la procédure PSI («Personalized System of Instruction») de Keller permet aux étudiants de progresser à leur propre rythme.

Longtemps préconisée au Québec et au Canada, l'approche instrumentée, originaire des États-Unis, est fondée sur une formulation très spécifique des objectifs d'apprentissage. Ceux-ci structurent les interventions et les orientent vers la maîtrise du plus grand nombre d'objectifs par le plus grand nombre d'élèves. Cette pédagogie de maîtrise entraîne des limites pratiques et méthodologiques qui peuvent être dépassées par l'introduction de paradigmes mixtes qui allient la recherche d'informations q .alitatives à celles quantitatives qui proviennent d'instruments de mesure (Dassa, 1990). Qui plus est, les difficultés d'objectiver le processus de régulation «mettent en doute l'idée d'une gestion théorisée de l'enseignement, d'une ingénierie de type scientifique, qui permettrait de maîtriser les situations d'apprentissage avec autant d'assurance que la construction des ponts parce que chaque élément de la situation pourrait être analysé et contrôlé pour lui-même» (Cardinet, 1988, 168). Ces difficultés d'ordre épistémologique inhérentes à la pédagogie de maîtrise et, notamment, à son élément central, la régulation, ont été à l'origine de diverses tentatives d'élargissement de l'évaluation formative. Originaire de l'Europe francophone, ce mouvement a été repris au Canada francophone et, notamment, par divers chercheurs québécois[5]. Situé entre une théorisation trop rigide coupée de la problématique de l'efficacité et une approche technicisante coupée de la didactique et de la psychopédagogie, cet élargissement de l'évaluation formative se veut suffisamment pragmatique pour permettre la recherche de l'optimisation des apprentissages tout en rendant possible une redéfinition de la régulation. Celle-ci se doit d'être ouverte à l'intégration de la didactique et des nouvelles technologies à l'évaluation. Cependant, pour être efficace en contexte scolaire, la régulation doit être fondée sur les pratiques de l'évaluation en classe.

4. VERS UN CADRE CONCEPTUEL DES PRATIQUES DE L'ÉVALUATION

L'étude des effets des pratiques évaluatives dans le contexte des apprentissages scolaires a fait l'objet d'un grand nombre d'études. Il est à présent bien établi que ces pratiques ont un impact important aussi bien directement sur le rendement scolaire lui-même que sur les variables

5 À cet effet, soulignons la contribution de la revue *Mesure et évaluation en éducation* qui sert de forum conjoint aux chercheurs et aux spécialistes francophones du Canada et d'Europe.

affectives qui influencent indirectement les apprentissages. Selon Crooks (1988), l'évaluation faite dans le contexte de la classe a un impact direct et indirect prononcé qui peut être soit positif soit négatif. Il est donc essentiel de procéder à des études systématiques sur les pratiques évaluatives. La recherche dans ce domaine est d'autant plus nécessaire que les régimes pédagogiques actuels attribuent un rôle central à l'évaluation dans la gestion pédagogique des apprentissages ainsi qu'un rôle prépondérant à l'enseignant en tant que premier responsable de cette évaluation.

Dans ce contexte, l'essentiel des fonctions évaluatives des enseignants peut se définir comme étant l'ensemble des processus et des procedures qui permettent d'apprécier le niveau d'atteinte des objectifs pédagogiques par l'élève ainsi que de juger de l'efficacité de l'enseignement dans une perspective opérationnelle de prise de décision relative à la planification du cheminement de l'élève. Lorsque ces fonctions s'exercent en classe, l'évaluation s'élargit naturellement de sa définition habituelle de jugement de valeur fondé sur une mesure formelle (test, examen) à une acception générale qui inclut les informations qualitatives informelles (interrogation orale, observation spontanée, ...). On retrouve ici la notion d'appréciation des apprentissages, notion de plus en plus mise de l'avant dans le contexte de la classe (Stiggins, 1988).

En amont de l'évaluation, au niveau des processus, il est important de tenir compte de l'effet des variables affectives des enseignants, telles leurs systèmes de valeur, de croyances et d'attitudes, sur leurs pratiques évaluatives. On peut définir ces dernières comme étant l'ensemble des interactions entre l'enseignant et l'élève. Elles se présentent sous forme de communication orale, écrite ou gestuelle ayant pour but de recueillir l'information nécessaire au cheminement ultérieur de l'élève. Cet échange interactif doit donc s'accompagner d'informations connexes sur les variables affectives (concept de soi, croyances et attitudes) des élèves. Plus en aval, lorsque les processus s'expriment en procédures observables, deux types de comportement se prêtent à l'observation : ce que les enseignants font en classe et la description qu'ils en font, c'est-à-dire ce qu'ils disent qu'ils font.

Ainsi élargi, le domaine de référence de ces pratiques évaluatives peut se subdiviser en huit catégories générales : a) le but de l'évaluation;

b) les moyens de l'évaluation; c) le contexte de l'évaluation; d) l'utilisation des moyens d'évaluation incluant la préparation, l'administration et la correction des instruments de mesure ainsi que le déroulement des moyens de colliger des informations qualitatives informelles; e) les analyses quantitatives et qualitatives; f) les moyens d'informer les élèves, les parents et les divers agents éducatifs; g) les caractéristiques relatives à la qualité de l'évaluation; h) les aspects réglementaires de l'évaluation.

Ce domaine de référence peut servir directement de base à l'élaboration d'un questionnaire destiné aux enseignants relatif à leurs pratiques évaluatives. Il est cependant évident que ce type d'information doit être complété d'une grille d'observation de certaines de ces pratiques en classe. Les contraintes méthodologiques spécifiques à ce type d'instrument de mesure nécessitent une structuration propre à l'environnement de la classe. Il s'ensuit une spécification du domaine de référence décrit ci-haut qui doit nécessairement porter sur les catégories suivantes :

a) le contexte pédagogique général des actes évaluatifs réalisés en classe;

b) le contexte pédagogique général du matériel évaluatif utilisé en classe;

c) le contexte évaluatif général, soit l'obtention de l'information (examens, tests, ...), l'utilisation de l'information (discussions, échanges, ...) et la contribution de l'information (cours théorique, explications, ...);

d) la structure générale de la leçon en trois segments : introduction, déroulement et conclusion;

e) les caractéristiques de la communication verbale non prescriptive de l'enseignant;

f) les caractéristiques de la communication verbale prescriptive de l'enseignant;

g) les caractéristiques des comportements à incidences affectives de l'enseignant;

h) les caractéristiques des échanges entre l'enseignant et les élèves.

Le domaine de référence des pratiques évaluatives proposées ainsi que les spécifications propres à l'observation directe des actes évaluatifs en classe pourraient, par les instruments de mesure qu'ils permet-

traient de développer, servir de base à un modèle général de gestion de l'information des pratiques évaluatives dans un but de prise de décision de l'enseignant vis-à-vis son propre enseignement et des divers agents éducatifs par rapport à l'encadrement administratif et pédagogique des enseignants en ce qui concerne les modes de fonctionnement qui ont des incidences pédagogiques.

Le but ultime de telles utilisations serait d'ajuster les pratiques pédagogiques et organisationnelles qui s'inscrivent dans le cadre des contraintes institutionnelles de façon à améliorer l'intégration de l'évaluation à l'enseignement. C'est pourquoi plusieurs recherches québécoises ont tenté d'apprécier aussi bien les pratiques d'évaluation que les croyances et les attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves.

5. Les pratiques évaluatives : entre le projet et les contraintes

Des études récentes (Burelle, Gadbois, Parent et Séguin, 1991; Gadbois, Burelle, Parent et Séguin, 1991) démontrent que les enseignants de la région métropolitaine de Montréal, tout en comprenant et en acceptant le bien-fondé de l'évaluation formative se gardent bien de prendre toutes les dispositions nécessaires pour l'intégrer à leurs pratiques quotidiennes. «Il semble [...] régner chez les enseignants une discordance entre la théorie et la pratique concernant l'évaluation formative. En effet, ils valorisent l'évaluation formative, mais considèrent que sa pratique est irréaliste dans le contexte scolaire actuel[6]». Il existe donc une différence entre le concept théorique accepté et la volonté de le mettre en application. Cette résistance au changement pourrait s'expliquer par divers facteurs qui touchent en particulier à la charge d'enseignement ainsi qu'à une formation lacunaire, notamment en mesure et évaluation.

6 Résultats d'une enquête menée auprès de 200 enseignants de l'ordre primaire de la région de Montréal. Voir, à ce sujet, Parent, C., Séguin, S.P., Burelle, R. et Gadbois, L. (1991). Croyances et attitudes des enseignants de l'ordre primaire de la région administrative du Montréal métropolitain à l'égard de la pratique de l'évaluation formative des apprentissages de leurs élèves, *in* Actes du colloque international de l'Association internationale de la recherche scientifique en éducation, Carcassonne, France.

Les travaux de Dassa et Dumoulin (1990) et de Dumoulin (1991) portent sur les pratiques d'une centaine d'enseignants de mathématiques et de sciences du secondaire de la région de Montréal. Ils visent à établir un profil détaillé de l'évaluation des apprentissages des élèves telle qu'elle est pratiquée par les enseignants. Ces études établissent que la grande majorité (83 %) des enseignants échantillonnés n'ont jamais suivi de cours de base de mesure et d'évaluation des apprentissages. Qui plus est, le quart d'entre eux n'a jamais reçu de perfectionnement dans ce domaine. Cette lacune sur le plan de la formation pourrait expliquer pourquoi les enseignants du secondaire hésitent à construire eux-mêmes les instruments de mesure requis. De plus, selon Dumoulin, les enseignants ont relativement peu de temps pour s'occuper des difficultés individuelles. Le type d'interaction le plus souvent pratiqué par les enseignants reflète une préoccupation d'évaluation collective de la classe qui laisse peu de place aux fonctions diagnostique et individuelle. Dans ce contexte, il semble donc tout aussi difficile d'intégrer l'évaluation formative instrumentée et systématique en classe que l'évaluation diagnostique reliée aux apprentissages individuels des élèves.

Dans une série d'études portant sur les facteurs qui déterminent le diagnostic pédagogique[7], une grille d'observation (Fontaine, 1991) a été élaborée et validée selon le domaine de référence proposé ci-haut. La grille a servi à observer 21 enseignants de la région de Montréal des ordres primaire, secondaire et collégial pendant 107 séances d'une durée de 15 minutes chacune. Le temps moyen consacré à manifester des comportements qui visent à orienter l'élève dans son processus d'apprentissage, c'est-à-dire à manifester des comportements qui relèvent en propre des fonctions évaluatives de l'enseignant, est de 46 %. La proportion du temps consacré à des actes évaluatifs tend à décliner lorsqu'on passe du primaire (63 %) et du premier cycle du secondaire (62 %) au deuxième cycle du secondaire (48 %) et au collégial (32 %). Finalement, soulignons que les enseignants qui ont suivi une formation en mesure et évaluation ont accompli, en moyenne, le double des actes évaluatifs accomplis par les enseignants qui n'ont aucune formation dans ce domaine.

7 Élaboration d'un modèle d'intégration du diagnostic pédagogique aux apprentissages scolaires. CRSH, subvention n° 410-89-1010, C. Dassa, chercheur principal; cochercheurs D. Ajar, R. Auger, C. Parent et S.P. Séguin.

Ces études permettent de dégager un profil type de l'enseignant tout acquis aux valeurs et aux principes de l'évaluation formative, mais inhibé dans son action par un manque de formation et d'outillage pédagogique. Qui plus est, par rapport à l'individualisation de la remédiation, l'enseignant est porté à attribuer à d'autres facteurs qu'à celui de son propre enseignement les causes des difficultés de ses élèves. Les lacunes relatives aux préalables, les problèmes familiaux et les troubles de comportement semblent constituer un paravent moral qui lui permet de limiter ses interventions correctives. Habituellement, l'enseignant procède à une reprise collective des contenus qui présentent des difficultés pour l'ensemble de la classe. Plus rarement, lorsque l'organisation pédagogique de son école le prévoit, il consacre un nombre très limité de périodes (au maximum une heure par semaine) à des séances individuelles de récupération. Celles-ci sont restreintes aux quelques élèves qui présentent des difficultés sérieuses. Ce constat rejoint diverses analyses concernant la «pression à la différenciation» qui apparaît lorsque l'évaluation formative fait partie du projet éducatif mais qui, faute de moyens, ne peut être pratiquée avec constance. «Il (s'ensuit) des conflits et des frustrations, donc une régression à des méthodes d'enseignement et d'évaluation plus conformes aux contraintes des maîtres» (Perrenoud, 1988).

En bref, ces résultats québécois vont dans le sens d'autres travaux sur les pratiques d'évaluation réalisées au Canada (Rogers, 1990) et aux États-Unis (Stiggins, 1988; Stiggins et Bridgeford, 1985). Les enseignants passent une grande partie de leur temps à accomplir des actes évaluatifs en se basant surtout sur des moyens informels d'évaluer la performance en situation d'apprentissage («performance assessment»). Ils effectuent ces évaluations en s'aidant d'instruments qu'ils développent eux-mêmes du mieux qu'ils peuvent, alors que leur formation en mesure et en évaluation est souvent inexistante et qu'ils se trouvent dénués d'outils pédagogiques. Tous soulignent enfin le problème du manque de temps pour développer ou pour s'approprier un outillage de qualité et pour l'intégrer à leurs prestations.

6. LE BESOIN D'OUTILLAGE ET LES TECHNOLOGIES INFORMATIQUES

Dans ce qui précède, nous avons vu que la recherche de l'individualisation des apprentissages entraîne une pédagogie différenciée. Or, celle-ci ne peut exister dans un milieu aussi contraint que l'école actuelle et, *a fortiori*, être efficace sans un soutien qui réponde aux besoins décrits ci-dessus. À défaut d'un remaniement en profondeur des structures et de l'organisation scolaire, force est de regarder du côté des nouvelles technologies. Or, au Québec comme ailleurs, on se rend de plus en plus compte que les technologies informatiques offrent des possibilités pédagogiques qui méritent d'être explorées (cf. par exemple, Berthelot, 1987).

Il est évident que les technologies de l'information et, plus particulièrement, la micro-informatique sont susceptibles d'aider à la réalisation fonctionnelle de pratiques pédagogiques conformes aux principes de l'évaluation formative. Elles permettent, d'une part, de recueillir les données et de les analyser de façon cohérente et, d'autre part, un réinvestissement des résultats obtenus afin d'améliorer l'enseignement et l'apprentissage des élèves. Les diverses utilisations de ces technologies en éducation ou applications pédagogiques de l'ordinateur (APO[8]), rendent possibles le développement d'outils de support pédagogique axés sur les processus d'apprentissage (Conseil supérieur de l'éducation, 1990) et donc permettent de soutenir pleinement cette pédagogie différenciée dont on sait qu'elle est liée à l'évaluation formative (Huberman, 1988).

Trente années de recherches sur les APO établissent, et ceci de façon consistante, que celles-ci ont des effets positifs sur le rendement des étudiants (Fletcher, Hawley et Piele, 1990; Kulik, J.A. et Kulik, C.-L., 1983, 1985, 1986, 1987, 1991; Suppes, 1966). Cependant, il aura fallu que deux conditions soient remplies, à savoir la percée du

8 *Applications pédagogiques de l'ordinateur*, terme le plus fréquemment utilisé au Québec pour signifier tout usage des technologies informatiques à des fins d'enseignement, d'évaluation et d'apprentissage. Sont comprises dans cette appellation, entre autres, les applications évaluatives telles les examens et les diagnostics informatisés ainsi que les applications didactiques telles les tutoriels («intelligents» ou non), les exerciseurs, les simulations et les jeux.

marché scolaire, réalisée notamment à partir de l'introduction des micro-ordinateurs dans les années quatre-vingt; l'intégration de la technologie au projet éducatif, telle qu'elle est préconisée dans les programmes d'études officiels.

Cependant, l'impact des APO, comme celui de toute innovation, dépend d'une troisième condition, l'adoption de la technologie par les enseignants et son intégration à leurs pratiques; autrement dit, l'effet des APO repose sur son intégration à un contexte réel. En effet, l'introduction de l'innovation, qui implique un usage significatif, par exemple, dans le cadre d'un cours et non pas dans une situation «expérimentale» ponctuelle, doit être suivie d'un processus d'appropriation par les enseignants (Dassa, 1988). Il est bien connu qu'un certain nombre de facteurs jouent un rôle important dans cette appropriation (Chomienne, 1986; Chomienne et Vazquez-Abad, 1990). Parmi ceux-ci se trouvent, entre autres, la capacité qu'a l'enseignant de devenir autonome dans la conception d'interventions pédagogiques qui impliquent l'innovation, l'adéquation de l'innovation à sa propre pratique pédagogique et à son style personnel d'enseignement ainsi que le support qui lui est offert : soutien technique, appartenance à des réseaux d'information, appui politique.

Au Québec, ce support s'est manifesté sous différentes formes au cours de la dernière décennie. Les investissements majeurs du Gouvernement en matière d'équipement ont été suivis d'initiatives pour encourager, voire subventionner, la production de logiciels éducatifs par l'entreprise locale.

Quant aux effets de la micro-informatique scolaire sur les apprenants, les études menées établissent, de façon consistante, des résultats positifs sur le rendement des étudiants. Plus particulièrement, lorsqu'on compare l'enseignement traditionnel avec un enseignement qui utilise les APO, quelle que soit l'application, on trouve qu'avec ce dernier, les élèves apprennent plus, mieux et plus rapidement, qu'ils aiment plus leurs cours et qu'ils développent une meilleure attitude envers l'informatique (Kulik, J.A. et Kulik, C.-L., 1983, 1985, 1986, 1987, 1991).

7. L'IMPORTANCE D'UNE CONCEPTION PÉDAGOGIQUE DES APO

Ainsi donc, des applications traditionnelles telles les exerciseurs en mathématiques (Hativa, 1988; Middleton, Curran et Moscardini, 1989) tout comme des applications plus récentes, telles les laboratoires de sciences assistés par la micro-informatique (Tinker et Papert, 1989) semblent offrir un grand potentiel pour la réussite scolaire. Toutefois, il est vraisemblable que leurs effets sur l'apprentissage soient dus à l'organisation du contenu et à la planification des interventions didactiques exigées par le design des APO plutôt qu'à l'utilisation même de l'ordinateur (Clark, 1985; Viau et Stolovitch, 1987). Il est maintenant acquis que pour obtenir des effets positifs sur l'apprentissage des élèves, il faut compter sur des logiciels dont l'efficacité est assurée par une conception théoriquement fondée et pédagogiquement solide (Baird, 1989).

La conception pédagogique serait donc une variable qui a une plus grande influence sur les résultats scolaires que le fait d'utiliser une technologie particulière. Ainsi, le succès d'un exerciseur axé sur le rendement (Vazquez-Abad et LaFleur, 1990), dont la conception repose tant sur la didactique de la matière que sur des résultats de recherche en psychologie cognitive et en médias éducatifs, serait dû à ses assises conceptuelles plutôt qu'à son application à l'aide de l'ordinateur.

Il ressort donc que la présence de la micro-informatique en milieu scolaire rend possible le développement de systèmes conçus pour soutenir des pratiques d'enseignement qui favorisent la régulation des apprentissages, c'est-à-dire conformes aux principes de l'évaluation formative.

8. DIAGNOSTIC PÉDAGOGIQUE ET INFORMATISATION

Nous présenterons maintenant le cadre conceptuel général d'un modèle[9] fondé sur la définition opérationnelle du diagnostic pédagogique

9 Ce modèle a été élaboré dans le cadre de la recherche mentionnée dans la note n° 7. Pour plus de détails concernant le choix de la stratégie de testing, voir Dassa, C. et Auger, R. (1991). *Computerized interactive diagnosis in a classroom setting*. Texte de la communication présentée au XIX[e] congrès annuel de la Canadian Society for the Study of Education, Queen's University. Pour une illustration

proposée par Dassa (1988, 1990), une brève description d'un système de tests diagnostiques conçu selon ce modèle ainsi qu'une appréciation de son efficacité.

Partie intégrante de l'évaluation formative, le diagnostic pédagogique consiste à identifier l'erreur, c'est-à-dire dans le contexte scolaire actuel, ce que l'enseignant identifie comme erroné, à décrire la nature de l'erreur, à décrire le contexte pédagogique où l'erreur se produit, à tenter de relier la (les) cause(s) de l'erreur à un paradigme de l'apprentissage, à proposer des interventions correctrices qui s'accordent à la didactique et à la technologie utilisées et qui s'inscrivent éventuellement dans un cadre théorique préalablement déterminé. L'actualisation de tout modèle élaboré à partir de cette conception du diagnostic passe nécessairement par la voie informatique afin d'assurer une mesure des apprentissages authentiquement individualisée et interactive. La dynamique de développement du modèle est articulée autour d'une stratégie de recherche rétroactive qui impliquent plusieurs acteurs : les chercheurs qui assument le support théorique et technique en vue de rendre opérationnel le modèle à l'aide d'une série d'unités d'apprentissage, les enseignants qui appliquent et vivent cette réalité. Cette dynamique permet l'intégration d'un grand nombre de facteurs touchant le curriculum, le système scolaire, l'organisation de l'enseignement ainsi que les caractéristiques pertinentes des professeurs et des élèves.

Le système informatique développé à partir de ce modèle est conçu de façon à maximiser la cohérence et l'efficacité de l'information obtenue. La cohérence du diagnostic est assurée par une banque d'items qui possède une structure à trois dimensions; ces items portent respectivement sur le contexte diagnostique, sur le contenu notionnel et sur l'habileté cognitive. Cette structure permet une catégorisation hiérarchisée des items qui, lorsqu'elle est couplée à une sélection adaptative (fondée sur les réponses de l'élève), permet un diagnostic structurel qui identifie la nature de l'erreur observée selon la taxinomie diagnostique

---

du modèle voir Auger, R. et Dassa, C. (1991). L'intégration du diagnostic interactif à l'évaluation formative informatisée des apprentissages scolaires. *Actes du colloque international de l'Association francophone internationale de la recherche scientifique en éducation*, Carcassonne, France, 15 p.

utilisée. Il est important de souligner ici que le système a été validé en étroite collaboration avec des enseignants qui ont contribué à la conception, à la mise à l'essai et à l'évaluation du système pendant deux ans. Qui plus est, les leurres des items à choix multiples sont établis à partir d'une analyse approfondie des erreurs commises réalisée conjointement par les chercheurs et les enseignants. Le diagnostic causal qui en résulte permet de proposer à l'enseignant une explication psychopédagogique de l'erreur commise qui reflète, outre les facteurs cognitifs habituels, ses propres pratiques pédagogiques. En ce sens, le diagnostic causal qui résulte de l'administration des tests diagnostiques constitue un outil de gestion respectueux des contenus, des programmes d'études, mais aussi de l'écologie de la classe. Qui plus est, le double mode d'utilisation de ce système permet de l'adapter à l'écologie de l'école. En effet, ce système peut fonctionner en laboratoire où toute la classe passe un test informatisé adaptatif relatif à un ou plusieurs objectifs spécifiques en vue d'établir un profil global des difficultés individuelles et collectives des élèves, et en récupération où l'enseignant rencontre, en dehors de la classe, quelques élèves qui présentent des difficultés jugées importantes et où il procède à un diagnostic approfondi en plusieurs phases, selon la (ou les) dimension(s) diagnostique(s) qu'il choisit.

Dans un tel environnement, l'appréciation des apprentissages s'appuie aussi bien sur des informations de nature qualitative que fondées sur des mesures et s'intègre aux stratégies d'enseignement et aux processus d'apprentissages selon une logique rétroactive, interactive et proactive (Allal, 1988). Les aspects rétroactifs se retrouvent dans la démarche d'évaluation collective en mode de laboratoire. À cause des contraintes d'horaire et bien qu'ils soient immédiatement disponibles, les résultats obtenus au test sont interprétés et intégrés dans la démarche d'apprentissage par l'enseignant lors du cours suivant. Quant à l'évaluation interactive, elle est exercée au cours de séances individuelles en mode de récupération. Il s'agit de rencontres privilégiées entre l'enseignant et l'élève en difficulté, soutenue par des informations diagnostiques structurées qui résultent d'une stratégie de *testing* adaptatif (sur mesure) géré par micro-ordinateur. L'enseignant et l'élève entrent en interaction au moyen de fiches diagnostiques individuelles (produites immédiatement) qui synthétisent les résultats obtenus en un profil de difficultés

qui sert à guider la remédiation. Les profils diagnostiques collectifs et individuels permettent alors à l'enseignant d'établir une stratégie d'intervention pédagogique proactive.

Ce système a été expérimenté dans deux écoles de la région de Montréal; l'une privée et l'autre publique. En tout, quatre enseignants de mathématiques de deuxième année du secondaire ont administré chacun huit séances de testing reliées à autant d'objectifs intermédiaires. Les élèves participants (environ 150) étaient regroupés en cinq classes expérimentales. Cinq autres classes qui regroupaient 150 élèves ont servi à la modalité de contrôle. Cette première phase de validation du modèle a permis d'établir que la grandeur de l'effet (*effect size*) est de l'ordre de 0,7 (sigma). Cette première indication de l'efficacité est encourageante à la lumière de travaux semblables en évaluation formative (Slavin, 1987).

9. Le diagnostic informatisé facteur d'intégration

Présentement, le système repose sur une analyse psychopédagogique qui permet de structurer la banque d'items. La stratégie de sélection des items est alors basée sur cette structure de façon à optimiser le diagnostic structurel et causal. Cette voie est clairement distincte des tests adaptatifs fondés sur la «théorie des réponses aux items» (TRI) qui visent à estimer une habileté unidimensionnelle. Or, nous savons que la TRI mène à une prise d'information optimale dans un environnement (d'items) unidimensionnel. Une piste de développement intéressante consisterait alors à identifier des ensembles d'items relevant chacun d'un facteur unique situé à l'intérieur de la structure diagnostique décrite ci-haut. De cette façon, on pourrait toujours baser le diagnostic sur cette structure de base qui définirait des phases de testing hiérarchisées selon des aspects didactiques et pédagogiques pertinents; chaque phase comporterait des étapes de prise d'information fondées sur la TRI. Vue sous cet angle, la recherche de stratégies de testing diagnostique informatisé devient un véhicule possible de la recherche d'une régulation élargie et améliorée porteuse d'efficacité.

CONCLUSION

Une autre avenue d'élargissement de cette régulation passe par l'intégration du diagnostic à la didactique. On peut ainsi concevoir des systèmes informatisés de diagnostic et d'apprentissage authentiquement intégrés (Dassa et Vazquez-Abad, 1991; Dassa et Vazquez-Abad, 1992). De tels systèmes pourraient constituer des outils pédagogiques qui, tout en aidant l'enseignant à centrer ses prestations sur les objectifs visés, lui permettraient de pratiquer une évaluation formative authentiquement individualisée basée sur une régulation différenciée élargie à un enseignement également différencié.

RÉFÉRENCES

ALLAL, L. (1988).
Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive. *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise* (p. 86-126). Paris : Delachaux et Niestlé.

BAIRD, W.E. (1989).
Status of use : Microcomputers and science teaching. *In* J.D. Ellis (éd.), *1988 AETS Yearbook : Information Technology and Science Education* (p. 85-104). Columbus, OH : AETS and ERIC Clearinghouse.

BARNARD, C.I. (1938).
The Functions of the Executive. Cambridge (MA) : Harvard University Press.

BERTHELOT, J. (1987).
La micro-informatique dans les écoles québécoises : résultats d'une enquête. *Canadian Journal of Education, 12*(3), 369-385.

BRUNET, L., BRASSARD, A. et CORRIVEAU, L. (1991).
Administration scolaire et efficacité dans les organisations. Montréal : Agence d'Arc.

BURELLE, R., GADBOIS, L., PARENT, C. et SÉGUIN, S.P. (1991).
Mesurer des croyances et des attitudes pour mieux comprendre des comportements. *Mesure et évaluation en éducation, 14*(1), 61-77.

CAMERON, K.S. et WHETTEN, D.A. (1983).
Organizational effectiveness : One model or several? *In* K.S. Cameron et D.A. Whetten (éd.), *Organizational Effectiveness. A Comparison of Multiple Models* (p. 1-26). New York : Academic Press.

CARDINET, J. (1981).
L'évaluation formative, problème actuel. *In* Allal, L., Cardinet, J. et Perrenoud, P. (éd.), *L'évaluation formative dans un enseignement différencié* (p. 10-18). Berne : Peter Lang.

CARDINET, J. (1988).
La maîtrise, communication réussie. *In* M. Huberman (éd.), *Assurer la réussite des apprentissages scolaires? Les propositions de la pédagogie de maîtrise* (p. 155-195). Paris : Delachaux et Niestlé.

CHOMIENNE, M. (1986).
L'enracinement des applications pédagogiques de l'ordinateur : une étude de cas dans la région de Montréal. Thèse de doctorat non publiée, Université Concordia.

CHOMIENNE, M. et VAZQUEZ-ABAD, J. (1990).
L'émergence du concept d'enracinement des applications pédagogiques de l'ordinateur. *Revue des sciences de l'éducation, 16*(1), 91-104.

CLARK, R.E. (1985).
Confounding in educational computing research. *Journal of Educational Computing Research, 1*, 137-148.

CROOKS, T.J. (1988).
The impact of classroom. Evaluation practices on students. *Review of Educational Research, 58*(4), 438-481.

CONSEIL SUPÉRIEUR DE L'ÉDUCATION (1990).
La pédagogie : un défi majeur de l'enseignement supérieur. Avis au ministre de l'Enseignement supérieur et de la Science. Québec : CSE.

CONSEIL SUPÉRIEUR DE L'ÉDUCATION (1987).
La qualité de l'éducation : un enjeu pour chaque établissement. *Rapport 1986-1987 sur l'état et les besoins de l'éducation.* Québec : CSE.

CUMMINGS, L.L. (1977).
Emergence of the instrumental organization. *In* Goodman, P.S. et Pennings, J.M. (éd.), *New Perspectives on Organizational Effectiveness* (p. 56-62). San Francisco : Jossey-Bass.

DASSA, C. (1988).
L'intégration du diagnostic pédagogique aux apprentissages scolaires : de la théorie à la voie informatique. *Mesure et évaluation en éducation, 11*(1), 7-26.

DASSA, C. (1990).
From a horizontal to a vertical method of integrating educational diagnosis with classroom assessment. *The Alberta Journal of Educational Research, 36*(1), 35-44.

DASSA C. et DUMOULIN, J. (1990).
Les pratiques de l'évaluation dans le contexte de la classe. Communication présentée dans le cadre du XIII[e] Colloque annuel de l'Association pour le développement de la mesure et de l'évaluation en éducation, Montréal.

DASSA, C. et VAZQUEZ-ABAD, J. (1991).
Computer based testing and drill and practice at a crossroads : Converging or diverging? *In* Winterburn, R. (éd.), *Realizing Human Potential* (p. 192-196). Londres : Kogan Page.

DASSA, C. et VAZQUEZ-ABAD, J. (1992).
De l'évaluation informatisée à l'intervention pédagogique. *Mesure et évaluation en éducation, 15*(3), 17-24..

DUMOULIN, J. (1991).
Étude des pratiques de l'évaluation des apprentissages réalisées par les enseignants de mathématiques et de sciences au secondaire. Mémoire de maîtrise non publié. Faculté des études supérieures, Université de Montréal.

FLETCHER, J.D., HAWLEY, D.E. et PIELE, P.K. (1990).
Costs, effects, and utility of computer assisted instruction in the classroom. *American Educational Research Journal, 27*(4), 783-806.

FONTAINE, S. (1991).
Élaboration et validation d'une grille d'observation visant l'identification des actes évaluatifs des enseignants dans un contexte d'évaluation formative. Mémoire de maîtrise non publié, Université de Montréal.

GADBOIS, L., BURELLE, R., PARENT, C. et SÉGUIN, S.P. (1991).
Un instrument de mesure des croyances et des attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves. *Mesure et évaluation en éducation, 14*(2), 5-24.

GOLDSTEIN, H. (1991).
Recontextualising Mental Measurement. Texte de la communication présentée au Symposium international sur les théories modernes de la mesure : nouveaux enjeux et nouveaux problèmes. Montebello (Québec) : Université d'Ottawa.

HATIVA, N. (1988).
Computer-based drill and practice in arithmetic. *American Educational Research Journal, 25*, 366-397.

HUBERMAN, M. (éd.) (1988).
Maîtriser les processus d'apprentissage. Fondements et perspectives de la pédagogie de maîtrise. Paris : Delachaux et Niestlé.

KANTER R.M. et BRINKERHOFF, D. (1981).
Organizational performance : Recent developments in measurement. *Annual Review of Sociology, 7*, 321-349.

KEELEY, M. (1984).
Impartiality and participant theories of organizational effectiveness. *Administrative Science Quarterly, 29*(1), 1-25.

KULIK, C.-L., KULIK, J.A. et BANGERT-DROWNS, R.L. (1990).
Effectiveness of mastery learning programs : A meta-analysis. *Review of Educational Research, 60*(2), 265-299.

KULIK, C.-L. et KULIK, J.A. (1991).
Effectiveness of computer-based instruction : An updated analysis. *Computers in Human Behavior*, (1/2), 75-94.

KULIK, J.A. et KULIK, C.-L. (1987).
Review of recent research literature on computer-based instruction. *Contemporary Educational Psychology, 12*(3), 222-230.

KULIK, C.-L. et KULIK, J.A. (1986).
Effectiveness of Computer-Based Education in Colleges. *AEDS Journal, 19*(2/3), 81-108.

KULIK, J.A. et KULIK, C.-L. (1985).
Effectiveness of computer-based education in elementary schools. *Computers in Human Behavior, 1*(1), 59-74.

KULIK, J.A. et Kulik, C.-L. (1983).
Effects of computer-based teaching on secondary school students. *Journal of Educational Psychology, 75*(1), 19-26.

MIDDLETON, W., CURRAN, D.A.S. et MOSCARDINI, A.O. (1989).
The use of computers in teaching mathematics. *Educational and Training Technology International, 26-4*, 310-317

MINISTÈRE DE L'ÉDUCATION DU QUÉBEC (1979).
L'École québécoise. Énoncé de politique et plan d'action. Québec : Ministère de l'Éducation du Québec.

MORIN, E.M. (1989).
Vers une mesure de l'efficacité organisationnelle : exploration conceptuelle et empirique des représentations. Thèse de doctorat non publiée, Faculté des études supérieures, Université de Montréal.

PERRENOUD, P. (1991).
Pour une approche pragmatique de l'évaluation formative. *Mesure et évaluation en éducation, 13*(4), 49-81.

PERRENOUD, P. (1988).
Formation à l'évaluation : entre réalisme conservateur et idéalisme béat. *In* M. Gather Thurler et Perrenoud, P. (éd.), *Savoir évaluer pour mieux enseigner. Quelle formation des maîtres?* Cahier n° 26. Genève : Service de la recherche sociologique.

PFEFFER, J. et SALANCIK, G.R. (1978).
The External Control of Organizations. New York : Harper and Row.

ROGERS, W.T. (1990).
Educational Measurement in Canada. Evolution or extinction? Texte de la Communication de l'invité du Président (ACCÉ), XVII[e] Congrès annuel de la Canadian Society for the Study of Education, Victoria.

SLAVIN, R.E. (1990).
Mastery learning re-reconsidered. *Review of Educational Research, 60*(2), 300-302.

SLAVIN, R.E. (1987).
Mastery learning reconsidered. *Review of Educational Research, 57*(2), 175-213.

STIGGINS, R. J. (1988).
Revitalizing classroom assessment : The highest instructional priority. *Phi Delta Kappan, 65*, 363-368.

STIGGINS R.J. et BRIDGEFORD N. (1985).
The ecology of classroom assessment. *Journal of Educational Measurement, 22*(4), 271-286.

SUPPES, P. (1966).
The uses of computers in education. *Scientific American, 215*, 206-220.

TINKER, R.F. et PAPERT, S. (1989).
Tools for science education. *In* Ellis, J.D. (éd.). *1988 AETS Yearbook : Information Technology and Science Education* (p. 1-24). Columbus (OH) : AETS and ERIC Clearinghouse.

VAN DE VEN, A.H. et FERRY, D.L. (1980).
Measuring and Assessing Organizations. New York : John Wiley.

VAZQUEZ-ABAD, J. et LAFLEUR, M. (1990).
Design of a performance-responsive drill and practice algorithm for computer-based training. *Computers and Education, 14*(1), 43-52.

VIAU, R. et STOLOVITCH, H.D. (1987).
L'impact de l'ordinateur sur l'apprentissage : réalité ou illusion? Montréal : APO-Québec.

ZAMMUTO R.F. (1984).
A comparison of multiple constituency models of organizational effectiveness. *Academy of Management Review, 2*(4), 606-616.

UN TEST ADAPTATIF EN LANGUE SECONDE : LA PERCEPTION DES APPRENANTS[1]

MICHEL LAURIER
Université de Montréal

INTRODUCTION

Une des questions principales que soulève la médiatisation de l'enseignement a trait aux effets qu'elle peut entraîner dans le processus d'apprentissage. Les effets de la médiatisation sont attribuables non seulement à la transposition d'une activité pédagogique d'un environnement à un autre mais aussi à la transformation que subit une activité quand on cherche à tirer le meilleur parti des possibilités qu'offre la technologie retenue. Ainsi, l'informatisation d'un test peut impliquer que les items qui le composent se comportent différemment lorsque l'administration se fait à l'aide d'un ordinateur. Plus encore, il est probable que cette informatisation se justifiera par le fait qu'elle permet de modifier l'administration afin de l'individualiser et de la rendre plus efficace.

C'est ainsi qu'au déroulement linéaire d'un test conventionnel conduisant au calcul d'un score s'oppose l'approche adaptative d'un test informatisé où chaque réponse amène une réévaluation de la performance en vue de modeler l'instrument à l'élève. On peut alors se demander comment les élèves réagissent face à un test adaptatif informatisé par rapport au test conventionnel qui leur est plus familier. Dans le cadre de ce chapitre, nous analysons les réactions d'élèves face à des modes d'administration différents en présentant deux études comparatives. Toutefois, comme les résultats rapportés sont liés aux choix que nous avons faits dans l'élaboration de ces instruments de mesure en cause, nous décrirons, dans une première partie, la fonction de ces tests et la façon dont ils ont été mis au point.

---

1 Cette recherche a été rendue possible en partie grâce à une subvention provenant du fonds CAFIR de l'Université de Montréal.

1. LES INSTRUMENTS DE MESURE

1.1 Les objectifs du test

Le test vise à évaluer une compétence générale en français langue seconde. On s'intéresse plus particulièrement aux élèves qui s'inscrivent à des cours de français dans les établissements anglophones d'enseignement postsecondaire au Canada. Ces établissements sont souvent confrontés, au début d'un semestre, au problème complexe du classement des élèves en fonction de leur niveau réel dans la langue seconde. Souvent désuets, les instruments qu'utilisent les établissements ne cadrent plus avec les approches pédagogiques et les contenus des nouveaux programmes de formation linguistique. Qui plus est, des considérations pratiques (manque de temps et de ressources) conduisent parfois à réduire l'opération de classement à un tri rudimentaire basé sur des données limitées et souvent peu pertinentes. C'est dans ce contexte que nous avons entrepris de mettre au point un test de classement en français langue seconde à l'intention des apprenants de niveau postsecondaire. À cet impératif pratique, se superposait une perspective de recherche où l'on souhaitait comparer une version conventionnelle d'un test de classement avec une version adaptative informatisée. Il nous semblait en effet que la question du classement en langue seconde se prêtait bien à la problématique d'une telle comparaison. D'une part, comme le test de classement doit couvrir toute la gamme des niveaux (du «parfait débutant» jusqu'au «très avancé»), plusieurs items qu'on y retrouve ne correspondent pas au niveau de l'élève; ces items apportent peu de renseignements sur l'élève et peuvent même le perturber. D'autre part, classer les élèves signifie, pour la majorité des programmes de formation linguistique, qu'on les ordonne (du plus débutant au plus avancé) sur un axe commun qu'on considère comme la «compétence générale» dans la langue seconde.

Le concept de «compétence générale» a fait l'objet d'un débat vif et fondamental entre les tenants de l'idée d'une compétence générale dominante, un facteur principal fort (Oller, 1979, 1983), et ceux qui considèrent que la capacité de communiquer dans une langue seconde doit se diviser en plusieurs sous-compétences (Bachman, 1990; Canale, 1981). Il semble que la compétence communicative soit effectivement

multidimensionnelle mais que, comme le fait remarquer Carroll (1987), la nature intégrative des apprentissages linguistiques fait en sorte que les sous-compétences se développent concurremment et que chaque acquisition est susceptible de devenir un élément facilitateur et parfois nécessaire à des acquisitions qui relèvent d'un autre ordre. Des études sur la composante factorielle des tests de langue (Davidson, 1988; Henning, 1992) font ressortir un facteur principal qui, s'il est difficile à définir sur le plan de la théorie linguistique et psychopédagogique, n'en représente pas moins un trait latent. On peut penser que contrairement à des tests qui portent sur des habiletés spécifiques, un test de classement bien construit ne contrevient pas au postulat d'unidimensionnalité. La théorie des items, le cadre théorique sur lequel se fonde, sur le plan psychométrique, la plupart des tests adaptatifs qu'on connaît, semble ainsi convenir ainsi pour l'élaboration d'un test de classement en langue seconde.

Pour cette application, les contraintes qu'exercent les exigences de la théorie de réponse aux items ne se manifestent donc pas tant par le respect de l'unidimensionnalité que par les restrictions qu'imposent la correction dichotomique sur laquelle reposent les modèles habituels et la taille de l'échantillon requis pour une calibration juste des items. À ces contraintes théoriques, s'ajoutent des contraintes technologiques. L'émission d'un message via un clavier et la réception d'un message via un écran amènent des échanges qui paraissent assez éloignés de la situation naturelle d'utilisation d'une langue, particulièrement quand on les compare à un véritable dialogue. Cette limite technologique risquait, dans notre cas, d'affecter d'autant plus la validité de l'instrument que nous cherchions à minimiser l'usage du clavier tout en nous en tenant à une technologie simple, accessible et abordable : le test devait fonctionner avec les configurations de base des ordinateurs personnels de la famille IBM. De plus, nous avons renoncé à utiliser des types d'items qui auraient exploité certains avantages propres à la machine (son, animation, rétroaction...), car nous avions prévu calibrer les items en administrant une version expérimentale de façon conventionnelle. Compte tenu des fins de la recherche, nous voulions maintenir la comparabilité des versions.

1.2 Le contenu du test

Malgré ces contraintes théoriques et technologiques, nous avons pu construire un test de classement qui mesure effectivement la compétence générale que nous cherchions à mesurer. Nous avons pu vérifier la validité de l'instrument en examinant les corrélations avec d'autres types de mesure. Pour nous assurer de l'unidimensionnalité des banques d'items, nous avons choisi de regrouper les items dans trois sous-tests qui se distinguent principalement par la nature de la tâche qu'on propose à l'élève. Dans les trois cas, il s'agit de réponses à choix multiple.

Premier sous-test. — Compréhension globale. L'élève lit un court paragraphe (environ 30 mots) construit sur le modèle d'un texte authentique : par exemple, un certificat de garantie, une carte d'invitation, un extrait de journal... Il doit ensuite choisir l'énoncé qui répond à la question posée ou celui qui reformule le mieux le texte.

Deuxième sous-test. — Choix de l'énoncé approprié. L'élève lit, dans sa langue maternelle, la description d'une situation de communication : par exemple, une demande d'information, une prise de parole, un remerciement... Il doit alors choisir l'énoncé français qu'il juge le plus convenable compte tenu de la valeur sémantique des énoncés proposés et des aspects sociolinguistiques de la situation.

Troisième sous-test. — Phrases à trou. Sur le modèle du traditionnel exercice lacunaire, l'élève doit lire une phrase dans laquelle on a supprimé un mot. On lui demande de choisir le mot qui pourrait compléter la phrase. Les effacements correspondent habituellement à des difficultés de type lexical ou grammatical.

Il convient de relever, particulièrement pour les deux premiers sous-tests, qu'il s'agit de tests intégratifs. En effet, bien qu'il soit généralement possible de reconnaître une notion, une fonction ou un point de langue à évaluer, nous avons évité d'isoler ces éléments de sorte que leur maîtrise est mesurée en même temps que des difficultés concomitantes qui pourraient se trouver dans le contexte. Cette approche respecte les conditions réelles d'utilisation de la langue, car on ne vérifie pas des connaissances parcellaires mais la capacité de mettre en œuvre ces con-

naissances. La calibration des items assigne alors des paramètres non pas à un élément spécifique mais plutôt à la microtâche où peut apparaître cet élément.

1.3 La mise au point

Étant donné le type d'items, nous avons opté pour une calibration selon un modèle à trois paramètres (discrimination, difficulté et hasard). Pour ce faire, on a administré une version expérimentale de 150 items, 50 items par sous-test, à quelque 700 élèves. Certes, il s'agit d'un échantillon minimal pour une calibration à trois paramètres mais la qualité de l'échantillon a permis de réduire l'erreur d'estimation des paramètres. En effet, le groupe était hétérogène du point de vue du trait mesuré mais très homogène quant à sa composition sociodémographique. Tous les sujets étaient inscrits dans un établissement postsecondaire au Canada, ils maîtrisaient bien l'anglais et ils participaient, au moment de la cueillette des données, à une session intensive d'été en français. Ils provenaient habituellement d'un milieu plutôt favorisé et la plupart étaient âgés de 17 à 24 ans.

La calibration s'est faite à l'aide du logiciel MicroCAT (Assessment System Corp, 1987). On a rangé les items dans trois banques (une par sous-test) qui devraient s'élargir avec l'addition de nouveaux items au fil des années. On a programmé un système de gestion de la banque qui permettra ces additions ou des modifications aux items ou aux paramètres. Le système de gestion, écrit en Turbo-Pascal (Borland Int., 1985) pour le système d'exploitation DOS, permet aussi la simulation de tests adaptatifs selon deux algorithmes différents. L'algorithme que nous avons retenu pour l'administration informatisée dans le cadre de la présente recherche s'apparente à la procédure stradaptative développée par Vale et Weiss (1975). Nous avons divisé le continuum que représente la compétence générale à évaluer (*theta* dans la terminologie de la théorie de réponse aux items) en treize strates. Ces strates représentent les sept niveaux habituellement distingués :

Débutant	Intermédiaire faible	Avancé
Faux débutant	Intermédiaire moyen	Très avancé
	Intermédiaire fort	

À ces niveaux s'ajoutent six niveaux mitoyens (par exemple, «Intermédiaire faible +»). La cote «très avancé +» est réservée aux élèves forts qui présentent un score parfait. Pour chaque strate, on ordonne les items en fonction de la quantité d'information qu'ils fournissent pour cette strate.

Lors une séance de testing adaptatif, le programme informatique doit toujours présenter l'item dont le niveau correspond le mieux au niveau estimé. Dans le cadre d'une procédure stradaptative, il s'agit alors de déterminer à quelle strate correspond ce niveau et de faire apparaître à l'écran un nouvel item qui apportera le maximum d'information. La strate de départ est établie à partir de données préalables : l'auto-évaluation de l'élève et ses antécédents en français pour le premier sous-test, puis le résultat du test précédent pour les deux autres sous-tests. Dans la première partie de chaque sous-test, l'estimation du niveau est relativement simple : on passe à la strate suivante si l'item est réussi ou à la strate précédente s'il est ne l'est pas. Vers la fin du sous-test, on essaie de déterminer la strate en estimant le niveau selon une technique plus précise basée sur la recherche du maximum de vraisemblance. Le nombre d'items administrés dans un sous-test varie entre 7 et 12; entre ces extrêmes, c'est la marge d'erreur de l'estimation qui décide de l'interruption du sous-test. Après les trois sous-tests, on communique immédiatement le résultat à l'élève en lui indiquant à quel niveau général il se classe.

La banque d'items calibrés a aussi servi à composer deux formes parallèles d'une version conventionnelle «papier-crayon». Ces deux formes conventionnelles réunissent chacune 60 items (20 items par sous-test) dans un questionnaire. Bien que les items des deux formes soient différents, le parallélisme implique que les moyennes et les variances des deux formes sont égales. De plus, inspiré par la définition du parallélisme que donne Samejima (1977), nous avons veillé à ce que que les courbes d'information suivent un tracé similaire.

Dans le cas des deux formes de la version conventionnelle, les élèves lisent la question dans le questionnaire et inscrivent leurs choix de réponse sur une feuille de réponses. Dans cas de la version informatisée, l'élève tape la lettre qui correspond à son choix de réponse. La

façon de répondre est expliquée au début de l'épreuve, sur le questionnaire ou à l'écran. Chaque sous-test est précédé d'une explication de la consigne et d'un exemple.

2. La comparaison entre les versions

Comme les deux versions devraient conduire au même type de décision, on peut se demander si elles sont véritablement interchangeables. On souhaiterait par exemple choisir une ou l'autre version en fonction des contraintes institutionnelles : administration individuelle ou en groupe, disponibilité des appareils et du personnel, traitement des résultats... Wise et Plake (1989) signalent que la correspondance entre les deux modes d'administration en cause n'est pas toujours immédiate. Par ailleurs, il ne suffit pas de se pencher sur les propriétés psychométriques des instruments mais il convient aussi d'examiner la façon dont les principaux intéressés, les élèves eux-mêmes, les perçoivent. Il faut donc se demander quel impact le mode d'administration peut avoir sur les attitudes et les comportements pendant le test. Selon Madsen (1990), cet impact est mineur, même si les élèves semblaient avoir une attitude plus positive à l'égard d'un test de langue adaptatif administré par ordinateur qu'à l'égard d'un test conventionnel. Par ailleurs, il est clair que les plans psychométrique et psychologique sont interreliés et que les effets psychoaffectifs peuvent éventuellement avoir des incidences sur la qualité de la mesure. Par exemple, d'après certaines études (Johnson et Mihal, 1973; Pine *et al.*, 1979), il semblerait que dans le contexte scolaire américain, les élèves des minorités noires réussissent généralement mieux à un test adaptatif administré par ordinateur qu'à son équivalent «papier-crayon». Gordon (1992) remet en cause la pertinence des questions à choix multiple dans des tests de langue seconde parce que ce type de question fait diminuer la confiance en soi de certains élèves et affecte leur résultat. Pourtant, ce qui nous intéresse ici ce ne sont pas tant les résultats en tant que tels mais plutôt les réactions des élèves, c'est-à-dire les sentiments qu'éprouvent ces élèves face à l'instrument, les perceptions qu'ils en ont et les comportements qu'ils adoptent au cours du test. Afin de mieux cerner la dimension psychologique du test de classement que nous avons mis au point, nous avons mené deux études où on sollicitait les réactions des élèves après la passation d'une version du test.

2.1 Première étude

2.1.1 Méthodologie

La première étude a été réalisée auprès d'étudiants inscrits dans une session intensive de français de six semaines. Le programme était organisé par l'Université York de Toronto mais il se déroulait dans le milieu francophone de Saint-Georges de Beauce. Presque tous les participants bénéficiaient d'une bourse du Secrétariat d'État dans le cadre d'un programme fédéral d'apprentissage de la langue seconde. Comme ce programme exige des organisateurs locaux qu'ils administrent un test de classement au début et à la fin de la session, nous avons pu faire en sorte que chaque élève reçoive les deux versions du test. Environ la moitié des sujets ont alors fait la version adaptative informatisée comme prétest tandis que l'autre moitié a fait une des formes de la version conventionnelle. Lors du post-test, nous avons inversé les groupes de sorte que ceux qui avaient fait la version informatisée ont fait la version conventionnelle et vice-versa. Les tests conventionnels ont été administrés en une seule séance pour tout le groupe alors que le test informatisé se déroulait dans un laboratoire de micro-informatique équipé d'une quinzaine de postes de travail. Immédiatement après le test, les élèves devaient remplir un court questionnaire. Comme 88 élèves étaient présents, à la fois pour le prétest et pour le post-test, on a donc recueilli 176 questionnaires.

Le questionnaire se distingue des questionnaires utilisés habituellement pour mesurer les attitudes par rapport à un test (notamment Shohamy, 1982; Scott, 1986; Zeidner et Bensoussan, 1988; pour les épreuves d'expression orale en langue seconde). En effet, ces derniers questionnaires sont construits à partir d'échelles Likert uniformes; ils sont conçus avec l'intention de soumettre les réponses à une analyse factorielle en vue de réduire les données à un nombre minimal de facteurs. Ils comportent donc certaines redondances et visent le plus souvent à porter un jugement global sur l'attitude des apprenants relativement à un mode d'évaluation particulier.

Notre approche est quelque peu différente, car nous avons plutôt cherché à confirmer un certain nombre d'hypothèses quant à la réaction

des élèves vis-à-vis des versions conventionnelles et informatisées de notre test. Chaque question était destinée à être analysée individuellement et porte sur un aspect particulier. Le questionnaire comportait un nombre limité questions, car il nous semblait que la concision était essentielle pour obtenir des données valables en demandant à des élèves de remplir un questionnaire immédiatement après un test. Comme nous voulions fournir au répondant une gamme d'options significatives, le nombre de possibilités de réponses pouvait varier. Bien qu'il ne se prêtait pas à une analyse statistique très fine, ce questionnaire nous semblait approprié pour mettre le doigt sur des contrastes dominants entre les deux types de tests.

2.1.2 Résultats

Les premières questions portaient sur certains attributs sociodémographiques du groupe. On a ainsi appris que les deux tiers des participants étaient inscrits dans des programmes universitaires appartenant aux humanités ou aux sciences sociales, la majorité d'entre eux se spécialisant d'ailleurs en français. L'autre tiers était constitué d'étudiants en administration, en éducation, en sciences/mathématiques ou aux beaux-arts. La moyenne d'âge était de 23 ans. Même si tous maîtrisaient bien l'anglais, pour le quart d'entre eux, il s'agissait d'une deuxième langue. La plupart disaient utiliser un ordinateur à l'occasion; moins de 3 % n'avaient jamais touché ce genre d'appareil alors que seulement 15 % passaient plus de trois heures par semaine devant un ordinateur.

Huit questions portaient sur les réactions face au test. Le tableau de la page suivante présente le sommaire des réponses obtenues.

La difficulté. — Comme le test adaptatif tient compte du niveau d'habileté du sujet, on s'attendait à ce que les élèves ne le trouvent ni difficile, ni facile. Par contre, les élèves avancés auraient dû percevoir le test conventionnel comme trop facile alors que les débutants auraient dû le percevoir comme trop difficile. On a donc demandé aux élèves de dire s'ils avaient trouvé le test facile ou difficile. En assignant une valeur numérique aux cinq choix de réponse proposées (de 1, pour «très facile»

Tableau 1 – Répartition des réponses au questionnaire

Questions	Réponses	Papier-crayon	Info. adaptatif	Total (N = 176)
Comment évaluez-vous le degré de difficulté du test?	Très facile	3	3	3
	Facile	17	20	37
	Juste	47	50	97
	Difficile	18	14	32
	Très difficile	3	1	4
Si ce test devait être utilisé pour vous classer, selon vous, où auriez-vous été classé?	Au-dessus de mon niveau	13	9	22
	Sous mon niveau	8	11	19
	Au bon niveau	67	67	134
Comment ce type de test mesure-t-il votre niveau général en français?	Avec précision	5	7	12
	Assez bien	70	73	143
	Mal	12	6	18
Que pensez-vous de la durée du test?	Trop long	10	2	12
	Trop court	5	10	15
	Bien	72	75	174
Les directives étaient-elles claires?	Très claires	72	78	150
	Généralement claires	3	1	4
	Ambiguës	13	9	22
Vous sentiez-vous à l'aise dans l'environnement du test?	Très détendu	31	42	73
	Détendu	54	40	94
	Tendu	3	6	9
	Très tendu	0	0	0
Vous était-il difficile de vous concentrer pendant le test?	Très difficile	1	1	2
	Difficile	17	19	36
	Facile	54	52	106
	Très facile	16	16	32
Que faisiez-vous le plus souvent quand vous ignoriez la bonne réponse?	Deviner en choisissant au hasard	63	60	123
	Deviner en répétant la même réponse	2	2	4
	Omettre la réponse	5	7	12

à 5, pour «très difficile»), on pouvait penser que les moyennes ne seraient pas sensiblement différentes mais que la variance différerait d'une version à l'autre. De fait, la moyenne des 88 sujets qui ont répondu au questionnaire après le test informatisé est de 2,89 et celle des 88 sujets qui ont répondu après le test conventionnel de 3,01. Une telle différence ne saurait être significative ($p < 0{,}294$). En examinant la distribution des réponses, on voit que ce léger écart tient au fait qu'un peu moins de sujets ont jugé la version informatisée «difficile» ou «très difficile». Une conclusion s'impose : bien que la procédure du test adaptatif consiste à présenter des items qui sont d'un niveau correspondant à celui du sujet, celui-ci n'a pas nécessairement l'impression de faire un test sur mesure. On peut donc penser que le niveau de difficulté objectif d'un test tel que traduit par les paramètres de difficulté ne correspond pas nécessairement à la difficulté subjective que perçoivent les sujets.

Le niveau. – Puisque les élèves ne jugent pas de façon différenciée le niveau de difficulté des deux versions, il ne faut pas s'étonner qu'ils portent le même jugement sur le niveau auquel ils ont été classés, quelle que soit la version utilisée. Ainsi quand on leur demande s'ils considèrent que le test les aurait placés au niveau approprié la plupart des élèves (76 %) estiment n'avoir été ni sous-évalués, ni surévalués. Ainsi, le fait que les items soient choisis en fonction du niveau estimé du sujet ne contribue pas à modifier la perception de ces sujets quant à l'exactitude de la décision relative à leur classement. Il est important de noter que les élèves qui venaient de faire la version informatisée avaient déjà pris connaissance du résultat final.

La précision. – Avec autant de sujets qui croient qu'ils auraient été placés à un niveau approprié grâce au test, on pourrait s'attendre à ce que la majorité considère que le test mesure leur compétence en français avec précision. Il n'en est rien puisque la grande majorité (83 %) juge simplement que le test mesure «assez bien». Par ailleurs, on aurait pu croire que le test conventionnel aurait été considéré comme plus précis du fait qu'il est plus long. Or, en donnant une valeur numérique aux trois catégories («bien» = 1, «assez bien» = 2, «mal» = 3), on constate que la différence entre les moyennes n'est pas significative. Qui plus est, 12 sujets (deux fois plus que la version informatisée) pensent que le test conventionnel mesure mal le niveau général en français.

Il est intéressant de noter que ce sont les sujets les plus jeunes qui semblent les plus critiques envers la version conventionnelle.

La longueur. – On demandait aux élèves de se prononcer sur la durée du test. On sait que la version conventionnelle dure environ une heure alors que la version informatisée dure environ 25 minutes. Il est donc surprenant que, malgré une différence appréciable entre la durée objective des deux types de tests, la grande majorité des élèves ne les jugent ni trop longs ni trop courts. On peut penser que bien que peu sensibles au fait que la procédure adaptative choisisse les items en fonction de leur niveau, les élèves reconnaissent qu'on ne peut apprécier la durée de deux types de tests avec les mêmes critères. On observe néanmoins qu'une dizaine de sujets jugent le test informatisé trop court et qu'un nombre égal juge le test conventionnel trop long. Il semblerait que plus les sujets sont familiarisés avec les ordinateurs, plus le test conventionnel leur paraît long.

La clarté. – Dans un test, l'élève doit savoir ce qu'on attend de lui. Au cours de la rédaction, nous avons porté un soin particulier à la clarté des directives. Interrogés à ce sujets, 72 sujets sur 88 jugent les directives de la version conventionnelle «très claires» alors que 78 sur 88 sujets portent un pareil jugement sur les directives de la version informatisée. Toutefois, la différence peut difficilement être considérée comme significative ($p < 0{,}263$). Par contre, il faut noter que de ce point de vue, les élèves les plus jeunes (25 ans et moins) ont tendance à juger plus favorablement la version informatisée alors que les plus âgés jugent la version conventionnelle plus positivement.

L'anxiété. – La situation de test et l'environnement informatique sont souvent des sources d'anxiété. Koch et Patience (1978) ont fait ressortir l'effet anxiogène des tests informatisés. Weiss et Betz (1976a, 1976b) ont trouvé que le test adaptatif informatisé jouait un rôle positif dans la motivation mais créait aussi de l'anxiété. Nous avons donc demandé aux participants s'ils se sentaient à l'aise dans l'environnement du test : classe, feuille de réponse, surveillant... par rapport à la salle d'informatique, clavier, écran... L'analyse des réponses montre que les élèves se disent très majoritairement «très détendus» ou «détendus» plutôt que «tendus» ou «très tendus». Si l'on assigne une valeur

numérique aux catégories de réponse (de 1 à 4), on ne trouve pas de différence vraiment significative entre les moyennes des deux versions (1,59 ~ 1,68). Par contre, la variance est différente de sorte qu'on peut croire que l'informatisation amène une plus grande polarisation : chez certains élèves la version informatisée cause plus d'anxiété alors que chez d'autres elle pourrait la réduire. Si on en juge par le résultat du test du chi carré ($p < 0{,}093$), cette répartition risque peu d'être le simple fruit du hasard. Elle ne semble pas lié à l'âge, ni à la familiarité avec les ordinateurs. On constate cependant que 50 % des sujets qui se spécialisent en français affirment être «très détendus» au test conventionnel par rapport à un peu moins de 45 % au test informatisé. En revanche, seulement 28 % des non-spécialistes se croient «très détendus» avec la version conventionnelle et près de 50 % avec la version informatisée. En d'autres termes, l'utilisation d'une version informatisée adaptative avec les non-spécialistes permettrait d'atteindre le niveau minimal d'anxiété que les spécialistes, plus habitués aux épreuves traditionnelles en français, atteignent avec la version «papier-crayon».

La concentration. – Nous nous intéressions à savoir si les élèves avaient éprouvé de la difficulté à se concentrer puisque l'environnement peut avoir des effets importants sur la capacité de concentration et conséquemment sur les résultats au test. La plupart des sujets (78 %) n'ont pas connu pas de problèmes de concentration. D'autre part, la cause des problèmes de ceux qui en éprouvaient ne réside sûrement pas dans le mode d'administration puisque les écarts entre les versions sont à peu près nuls.

Les stratégies. – Nous avons aussi demandé aux élèves ce qu'ils faisaient le plus souvent quand ils ignoraient la bonne réponse. Nous pensions que le fait d'utiliser un médium différent pouvait affecter les stratégies de réponse dans les cas où l'élève ne savait pas la réponse. Par exemple, il semble plus facile d'omettre une réponse avec une feuille de réponses puisqu'il n'y a aucun geste à poser alors qu'avec le clavier, il faut appuyer sur la barre d'espacement; de plus, dans la consigne même du test adaptatif, on invitait l'élève à utiliser cette touche le moins possible. Par ailleurs, il peut sembler plus facile de répéter la réponse précédente avec l'ordinateur puisqu'il suffit de garder le doigt sur la même

touche. On doit pourtant reconnaître que ces hypothèses ne se sont pas vérifiées. Tant pour la version informatisée que pour la version conventionnelle, répondre au hasard est nettement la stratégie la plus populaire quand l'élève ignore la réponse. Il faut cependant souligner que le principal commentaire qu'ont exprimé les élèves au bas du questionnaire concernait précisément cet aspect. En effet, beaucoup se défendaient de «deviner» précisant qu'ils procédaient en évaluant la vraisemblance de chaque distracteur.

2.1.3 Discussion

Il faut rappeler que cette étude visait à déterminer s'il y avait des raisons d'ordre psychologique évidentes, c'est-à-dire des différences marquées entre les modes d'administration, pour rejeter toute comparaison entre les résultats obtenus avec des versions différentes. Il apparaît clairement que pour ce type de test, avec la population étudié, il n'y a pas de différence marquée dans les réactions des élèves selon qu'on utilise la version adaptable ou la version «papier-crayon». Cette conclusion est certainement encourageante pour le développement d'instruments informatisés puisque l'administration de tels tests ne semble pas générer autant d'anxiété que ce qu'on anticipait. Par contre, le fait que les élèves qui se spécialisent en français pourraient être légèrement favorisés à ce chapitre mériterait d'être exploré.

Par ailleurs, l'absence de différence significative remet en cause certains avantages qu'on attribue habituellement aux tests adaptatifs. On peut notamment mettre en doute l'idée que le niveau où un item donné apporte le plus d'information soit nécessairement le niveau où cet item est jugé le plus approprié par le sujet.

2.2 Deuxième étude

2.2.1 Méthodologie

Ericsson et Simon (1987), deux psychologues pionniers des techniques d'observation à l'aide de protocoles verbaux, distinguent les techniques introspective des techniques rétrospectives. En introspection, le sujet fait des commentaires «sur le champ», c'est-à-dire pendant qu'il accomplit la tâche (exercice de compréhension, test, rédaction...).

En rétrospection, le(s) sujet(s), avec l'aide discrète d'un animateur, tente(nt) de retracer «après coup» la démarche intellectuelle suivie. Cohen (1984, 1987) montre que les rapports verbaux, introspectifs ou rétrospectifs, peuvent s'avérer fort utiles pour étudier un test en langue seconde. Il rapporte plusieurs études où l'on demandait aux sujets de commenter leur démarche pendant qu'ils complétaient un test de closure ou pendant qu'ils répondaient à des questions à choix multiple. Quant à nous, nous avons opté pour une technique rétrospective relativement simple, la discussion de groupe. Il nous semblait que les rapports verbaux ainsi recueillis serviraient à compléter les données du questionnaire.

Deux groupes d'une dizaine d'élèves anglophones inscrits dans un programme intensif de l'Université du Québec de Trois-Rivières ont fait chacun une version du test (conventionnelle ou informatisée). Tous les niveaux, sauf les débutants absolus, étaient représentés. La composition sociodémographique des groupes était semblable à celle qu'on trouvait avec le groupe de la première étude. Immédiatement après l'épreuve, on amorçait une discussion de groupe sur le test lui-même. Le rôle de l'animateur se limitait essentiellement à éviter les digressions en ramenant la discussion autour des principaux aspects visés : difficulté du test, stratégies de réponse, nervosité... Il faut noter qu'à ce moment, le logiciel avait déjà informé les sujets du test adaptatif de leur niveau alors que la correction du test conventionnel s'est faite après la discussion. On a analysé l'enregistrement des discussions dont nous résumons ici le contenu.

2.2.2 Le test adaptatif

La discussion s'est déroulée au laboratoire même où les dix élèves venaient de faire le test. De façon générale, les élèves semblaient avoir porté beaucoup d'intérêt au test même si une participante le trouvait «impersonnel». Unanimement, les élèves ont d'abord reconnu qu'ils n'avaient eu aucun problème avec la consigne, peu importe leur degré de familiarisation avec la machine. De fait, les explications que fournit le logiciel ont rassuré les plus réticents de sorte que personne n'a dit avoir éprouvé de crainte ou d'anxiété au cours du test. Les participants ont souligné que la nervosité qui est associée habituellement à

un test tient à la signification que prend le test pour le dossier scolaire et non au mode d'administration.

Contrairement à ce qu'on pourrait attendre d'un test adaptatif, les participants ont indiqué que le niveau de difficulté de chaque item variait : certaines questions leur paraissaient nettement plus faciles ou plus difficiles que d'autres. Tous s'entendaient pour dire que la première partie (sous-test de compréhension globale) était la plus difficile à cause de la complexité du vocabulaire. Les élèves ont comparé le test avec le test qui leur avait été administré au début de la session comme test de classement. Celui-ci, selon les plus débutants, était beaucoup plus difficile alors que selon les plus avancés, il était plus facile. On reconnaît ici l'effet de la procédure adaptative du point de vue de la perception de la difficulté générale du test. Interrogés quant à la proportion de questions auxquelles ils pensaient avoir répondu correctement, les élèves s'entendaient pour dire que le taux se situait autour de 80 %. Une seule personne a noté que le test s'adaptait à son niveau; les autres ont manifesté une certaine surprise en apprenant qu'ils avaient tous fait des tests différents.

Tous ont apprécié de connaître leur résultat sitôt le test terminé. Il ont tous fait remarquer que le niveau qui leur avait été communiqué correspondait à ce qu'ils croyaient être leur niveau réel sauf pour ce qui est de la performance à l'oral.

Quand on leur a demandé s'ils auraient préféré faire la version conventionnelle, les élèves ont hésité. Plusieurs ont fait remarquer que les stratégies de test qu'ils utilisaient habituellement ne fonctionnaient pas puisqu'ils étaient dans l'impossibilité de faire de révisions. Ils ont noté que le mode de présentation empêchait l'utilisation d'éléments des autres questions comme indices. Par ailleurs, aucun élève n'aurait utilisé la barre d'espacement pour omettre une réponse. Quand ils ignoraient la réponse, disaient la plupart, ils tentaient de repérer les deux distracteurs les moins plausibles pour restreindre le choix à une simple alternative.

2.2.3 Le test conventionnel

Pendant que le groupe qui se trouvait au laboratoire de micro-informatique faisait le test adaptatif puis en discutait, un deuxième groupe de 11 élèves travaillait avec la version «papier-crayon». La dernière copie remise, on a entamé la discussion. D'emblée, tous ont admis qu'avec un test à choix multiple comme celui qu'ils venaient de faire, la tâche était très claire. Cependant, on a remis en question la précision des tests à choix multiple, d'autant plus qu'il n'y avait pas de partie orale. Malgré cette réserve quant au contenu, tous désiraient connaître leur résultat et ont par la suite attendu que la correction soit complétée.

Les participants n'ont pas trouvé le test trop long. Personne n'était nerveux et les participants ont rappelé que dans ce contexte, il y avait peu de raison de l'être parce que la nervosité est reliée aux conséquences du test. Une participante a mentionné toutefois qu'elle aurait été nerveuse si on lui avait proposé un test informatisé. La majorité des participants ont par contre déclaré que, par curiosité, ils auraient opté pour cette dernière version s'ils avaient eu le choix.

Tous ont affirmé que la première partie exigeait beaucoup de lecture et qu'à cause de la complexité du vocabulaire, elle était nettement plus difficile. Ils ont signalé que la deuxième partie, malgré les jugements extralinguistiques qu'elle impliquait, leur avait paru la plus intéressante. Par ailleurs, interrogés sur la proportion des items qu'ils croyaient avoir réussis, les participants ont donné des réponses très variées : trois d'entre eux estimaient leur taux de réussite à 50 %, tandis qu'une participante pensait n'avoir presque aucune réponse correcte et qu'une autre pensait avait trouvé presque toutes les réponses.

En ce qui a trait aux stratégies mises en œuvre, la technique d'élimination des deux distracteurs les moins plausibles a de nouveau été mentionnée. On a nuancé toutefois en ajoutant qu'on cherchait parfois des mots clés et des indices, y compris des indices fournis par d'autres items. Par ailleurs, quoiqu'ils soient conscients de la possibilité de faire des révisions ou des retours, tous les participants, excepté un seul (de niveau avancé), ont avoué ne pas avoir exploité cette possibilité.

2.2.4 Discussion

Les sujets paraissaient plus critiques envers la formule des choix multiples avec la version conventionnelle qu'avec la version informatisée pour laquelle ils imaginaient peut-être mal une autre type de questions. Il est manifeste qu'aucun des tests ne suscitait de nervosité ou d'anxiété et que la version informatisée éveillait davantage l'intérêt des sujets. Par ailleurs, les commentaires quant au contenu ne changaient guère. Tous étaient d'accord avec la nécessité d'ajouter une composante orale. Tous s'entendaient également sur le fait que la première partie était nettement plus difficile.

Ce dernier commentaire sur la difficulté du premier sous-test peut surprendre chez ceux qui ont fait la version adaptative. En effet, en principe, les items des trois sous-tests se réfèrent à une échelle commune et c'est le niveau du sujet qui détermine la difficulté des items. On voit donc que la difficulté subjective des items ne correspond pas à leur difficulté objective (le paramètre de difficulté). Mais comment expliquer que les élèves font des prévisions assez justes quant à la proportion des items qu'ils ont réussi? Une hypothèse de recherche s'ouvre : la perception de la difficulté d'un test, (et peut-être de toute tâche langagière) ne dépendrait pas tant de considérations probabilistes (les chances de réussir) que de considérations touchant la nature même de la tâche. Dans cette perspective, concevoir l'adaptabilité d'un test à partir uniquement des paramètres de chaque item ne suffit plus pour en arriver à un test sur mesure.

Il est également intéressant de noter qu'on reproche au test informatisé d'empêcher les retours et les révisions puisqu'une fois qu'on a appuyé sur une touche l'item est irrécupérable. Même si beaucoup d'élèves disent ne pas avoir fait de retours ou de révisions de manière systématique, ils signalent que la recherche d'indices à travers tout le test fait partie des stratégies dont ils disposent. Ainsi, bien qu'on puisse prétendre satisfaire l'exigence d'indépendance des items avec la version informatisée, on ne saurait en dire autant de la version conventionnelle. Cela devient problématique quand on sait que la calibration s'est effectuée à partir d'un test «papier-crayon».

Enfin, en commentant leurs stratégies de réponse, les sujets sont unanimes pour dire qu'il n'ont jamais répondu tout à fait au hasard mais qu'ils préfèrent procéder par inférence («*educated guess*»). Quel est donc le rôle du troisième paramètre dans une analyse selon un modèle à trois paramètres si personne ne devine vraiment? Le nombre de sujets supplémentaires que requiert l'estimation de ce paramètre à la suite de l'expérimentation est-il justifié? On peut se demander, si le hasard ne joue pas, s'il serait mieux ne pas prendre en considération ce paramètre.

CONCLUSION

Les impressions et les commentaires que les élèves ont livrés à travers, d'une part, les questionnaires qu'ils ont remplis et, d'autre part, la discussion de groupe, montrent que l'informatisation d'un test de classement peut se faire sans que ne se modifient de façon marquée les attitudes ni les comportements des élèves. Dans le cas du test de classement que nous avons mis au point, les effets négatifs qu'on pouvait craindre, notamment en ce qui a trait à l'anxiété, semblent négligeables.

Par contre, l'application des principes de testing adaptatif ne produit pas l'effet escompté puisque les élèves sont assez peu sensibles à la correspondance qu'on recherche entre la difficulté des items et le niveau d'habileté. La perception des apprenants et les stratégies qu'ils adoptent nous amènent à nous interroger sur l'applicabilité de la théorie du trait latent dans l'élaboration d'un test adaptatif en langue seconde. On peut penser qu'une théorie psychométrique qui ignore la nature des tâches s'avère peu efficace pour construire des représentations de la difficulté. Par ailleurs, il faut se rappeler qu'en dehors de l'application des principes de testing adaptatif, le test que nous proposions tirait peu avantage du médium informatique. La présente recherche nous rappelle qu'une exploitation même poussée des possibilités pédagogiques de l'ordinateur, tant pour l'évaluation que pour l'enseignement, n'a pas nécessairement de répercussions dans les perceptions des apprenants à qui elle s'adresse.

RÉFÉRENCES

ASSESSMENT SYSTEM CORP. (1987).
User's Manual for the MicroCAT Testing System (2e éd.). MN : St-Paul.

BACHMAN, L.F. (1990).
Fundamental Considerations in Language Testing. New York : Oxford University Press.

BORLAND INTERNATIONAL INC. (1985).
Turbo-Pascal, Version 4.0. California : Scotts Valley.

CANALE, M. (1981).
Communication : How to evaluate it? *Bulletin de l'ACLA, 3*(2), 77-94.

CARROLL, J.B. (1987).
Psychometric theory and language testing. *In* Grotjahn, R.C., Klein-Bradley, C. et Stevenson, D.K. (éd.), *Taking their Measure: The validity and validation of language tests* (p. 1-40). Bochum : Studienverlag Dr N. Brockmeyer.

COHEN, A.D. (1984).
On taking language tests : What the students report. *Language Testing, 1*(1), 70-81.

COHEN, A.D. (1987).
Using verbal reports in research on language learning. *In* Faerch, C. et Kasper, G. (éd.), *Introspection in Second Language Research* (p. 82-95). Clevedon : Multilingual Matters.

DAVIDSON, F.G. (1988).
An Exploratory Modeling Survey of the Trait Structures of Some Existing Language Test Datasets. Thèse de doctorat non publiée, Université de Californie à Los Angeles.

ERICSSON, K.A. et SIMON, H.A. (1987).
Verbal report on thinking. *In* Faerch, C. et Kasper, G. (éd.), *Introspection in Second Language Research* (p. 24-53). Clevedon : Multilingual Matters.

GORDON, C. (1992).
The effects of affective variables and testing methods on test performance in EFL. Communication au Language Testing Research Colloquium, Vancouver.

HENNING, G. (1992).
Dimensionality and construct validity of language tests. Communication au Language Testing Research Colloquium, Vancouver.

JOHNSON, D.F. et MIHAL, W.L. (1973).
Performance on Blacks and Whites in computerized versus manual testing environment. *American Psychologist, 28*(8), 694-699.

KOCH, B.R. et PATIENCE, W.M. (1978).
Students' attitudes toward tailored testing. *In* Weiss, D.J. (éd.), *Proceedings of the 1977 Conference on Computerized Adaptive Testing* (p. 116-127). Minneapolis (MN) : Université du Minnesota.

MADSEN, H. (1990).
Computer adaptive testing of listening and reading comprehension. *In* Dunkel, P. (éd.), *Computer-Assisted Language Learning and Testing : Research Issues and Practice* (p. 237-257). New York : Newbury House.

OLLER, J.W. JR. (1979).
Language Tests at School. Londres : Longman.

OLLER, J.W. JR. (1983).
Evidence for a general proficiency factor: an expectancy grammar. *In* Oller, J.W. Jr. (éd.), *Issues in Language Testing Research* (p. 3-10). Rowley (MA) : Newbury House.

PINE, S.M. *et al.* (1979).
Effects of Computerized Adaptive Testing on Black and White students. (Research Report 79-2). Minneapolis (MN) : Université du Minnesota, Département de psychologie.

SAMEJIMA, F. (1977).
A use of the information function in tailored testing. *Applied Psychological Measurement, 1*(2), 233-247.

SCOTT, M. (1986).
Students' affective reaction to oral language test. *Language Testing, 3*(1), 99-118.

SHOHAMY, E. (1982).
Affective considerations in language testing. *Modern Language Journal, 66*, 13-17.

VALE, C.D. et WEISS, D.J. (1975).
A Simulation Study of Stradaptive Ability Testing (Research report 75-6). Minneapolis (MN) : Université du Minnesota, Département de psychologie.

WEISS, D.J. et BETZ, N.E. (1976a).
Effects of Immediate Knowledge of Results and Adaptive Testing of Ability Test Performance (Research report 76-3). Minneapolis (MN) : Université du Minnesota, Département de psychologie.

WEISS, D.J. et BETZ, N.E. (1976b).
Psychological Effects of Immediate Knowledge of Results and Adaptive Testing (Research report 76-3). Minneapolis (MN) : Université du Minnesota, Département de psychologie.

WISE, S.L. et PLAKE B.S. (1989).
Research on the effects of administering tests via computers. *Educational Measurement Issues and Practices, 8*(3), 5-10.

ZEIDNER, M. et BENSOUSSAN, M. (1988).
College students attitude towards written versus oral tests of English as a foreign language. *Language Testing, 5*(1), 100-114.

ÉVALUER LA TRANSFORMATION DE STRUCTURES DE CONNAISSANCE PROPOSITIONNELLE CHEZ DES ADULTES EN FORMATION : QUESTIONS MÉTHODOLOGIQUES POUR LA RECHERCHE

ÉTIENNE BOURGEOIS
Université catholique de Louvain

INTRODUCTION

La question centrale que j'aborde ici est de savoir comment on peut identifier et évaluer les transformations cognitives qui s'opèrent chez des adultes en situation de formation, dans le domaine des connaissances propositionnelles. Cette problématique méthodologique est en effet au centre de nos préoccupations actuelles dans le cadre d'une lignée de recherche en cours de développement au sein de notre laboratoire.

Ces travaux portent sur les interactions entre, d'une part, le processus d'apprentissage chez des adultes en formation et, d'autre part, certaines variables liées aux caractéristiques de la situation de formation et des apprenants eux-mêmes. Leur objectif est d'identifier les facteurs susceptibles d'affecter significativement le processus d apprentissage dans ce contexte et de saisir les mécanismes par lesquels ils interagissent avec ce processus. Précisons encore que l'on s'intéresse ici plus particulièrement dans ce contexte aux apprentissages relatifs à des connaissances propositionnelles (formation de concepts et de théories) que l'on distingue, notamment, des connaissances procédurales. De plus, les connaissances propositionnelles étudiées dans ces travaux ont pour objet la réalité sociale. Actuellement, ces travaux se développent dans trois directions : a) les relations entre le processus d'apprentissage, le mode d'interaction sociale à l'œuvre au sein du groupe en formation et le mode de gestion de ces interactions par le formateur; b) les relations entre apprentissage, estime de soi et motivation; c) les relations entre apprentissage, projet de formation et trajectoire sociale de l'apprenant (Bourgeois et Nizet, 1989).

Dans une perspective de travail cumulatif, nous avons cherché à mettre au point un dispositif de recueil et de traitement des données pour appréhender de manière cohérente la variable apprentissage dans chacune de ces recherches. Avant d'évoquer plus en détail ce dispositif et certaines de ses variantes, j'en explicite les fondements théoriques principaux.

1. UNE PERSPECTIVE CONSTRUCTIVISTE DE L'APPRENTISSAGE

Dans les travaux susmentionnés, l'apprentissage est appréhendé dans une perspective constructiviste. Je ne vais pas redévelopper ici tous les principes essentiels qui caractérisent ce paradigme. Tardif en fait une brillante synthèse ailleurs dans cet ouvrage. Je me contenterai de rappeler brièvement deux idées centrales dans cette approche.

Tout d'abord, le paradigme constructiviste suppose que l'apprenant aborde des apprentissages nouveaux avec des structures de connaissance préalables. Apprendre, c'est précisément le processus par lequel ces structures de connaissance préalables se transforment en structures nouvelles au travers d'une confrontation dynamique à des informations nouvelles. On peut schématiser cela comme suit (cf. figure 1).

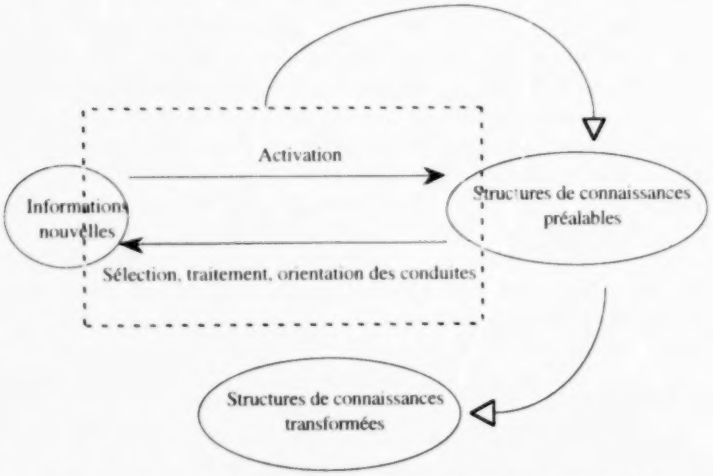


Figure 1 – L'apprentissage dans le paradigme constructiviste (formulation générale)

Dans cette perspective, étudier l'apprentissage suppose donc que l'on soit en mesure de décrire les états successifs d'une structure de connaissance donnée tout au long de sa transformation au cours de la séquence étudiée. Évaluer les résultats de l'apprentissage suppose que l'on puisse mettre en relation l'état de la structure au terme de la séquence avec le ou les états attendus, en fonction des objectifs d'apprentissage considérés.

La seconde idée vient compléter cette formulation générale du paradigme et a des implications méthodologiques fondamentales ainsi que nous le verrons. Certains psychologues cognitivistes (notamment, Richard, 1990; Cordier *et al.*, 1990) établissent en effet une distinction importante entre connaissances et représentations[1]. Le concept de représentation désigne chez eux la construction mentale d'une situation particulière, élaborée et mise en œuvre *hic et nunc* par le sujet confronté à cette situation. Il s'agit donc de

> [...] *constructions circonstancielles faites dans un contexte particulier et à des fins spécifiques, élaborées dans une situation donnée et pour faire face aux exigences de la tâche en cours un texte qu'on lit, une consigne qu'on écoute, un problème qu'on doit résoudre. Cette construction est finalisée par la tâche et la nature des décisions à prendre. Les représentations ainsi élaborées prennent en compte l'ensemble des éléments de la situation et de la tâche : elles sont de ce fait très particularisées et donc occasionnelles et précaires par nature. Il suffit que la situation change ou qu'un élément non remarqué de la situation soit pris en compte, alors qu'il ne l'était pas, pour que la représentation soit modifiée. Elles sont par nature transitoires : une fois la tâche terminée, elles sont remplacées par d'autres représentations liées à d'autres tâches* (Cordier *et al.*, 1990, 36).

Les représentations, en ce sens, se distinguent des connaissances. Celles-ci sont également des structures mentales mais contrairement aux représentations, elles sont stabilisées en mémoire à long terme et ne sont pas liées à des situations spécifiques.

1 Voir également Bourgeois (1992b) sur ce point.

Sur le plan du fonctionnement cognitif, les représentations sont étroitement liées aux connaissances. Prenons l'exemple d'une séquence d'apprentissage dans le cadre d'un cours de sociologie centré sur la notion de pouvoir dans les organisations. Au début de la séquence, l'enseignant présente, sous la forme d'une narration, une situation problème (un cas, un événement) qui met en jeu différents modes d'exercice du pouvoir entre des acteurs dans une organisation. Il demande ensuite à chacun des apprenants de formuler leur diagnostic du problème qui leur est soumis («comment interprétez-vous cette situation?») et de proposer des solutions («comment auriez-vous agi dans cette situation?»). Son objectif est de confronter par la suite les lectures spontanées de l'événement produites par les apprenants et la lecture que l'on peut en faire à la lumière de la théorie sociologique qu'il veut enseigner.

Dans cette situation de formation, l'apprenant est donc amené à produire une représentation de la situation-stimulus qui lui est présentée. Les éléments d'information descriptive qu'il retient de la situation-problème (concernant les acteurs en présence dans cette situation, leur comportement, leur interactions, etc.), les catégories conceptuelles qu'il utilise pour nommer et qualifier ces éléments, les différentes relations (par exemple, de cause à effet) qu'il établit entre ces éléments, les inférences qu'il produit à partir de ces éléments et de ces relations, les jugements qu'il porte sur ces éléments et leurs relations, les stratégies d'action qu'il propose en regard de cette situation, tout cela constitue la représentation qu'il se fait de la situation stimulus en question.

La production de cette représentation est cependant étroitement liée aux connaissances stockées en mémoire à long terme de l'apprenant et qui auront été activées par lui relativement à cette situation problème. En «amont», le sujet va construire sa représentation de la situation stimulus non pas *ex nihilo*, mais bien en s'appuyant notamment sur des structures de connaissance préalables, stockées en mémoire, et qu'il va activer pour faire face à la situation en question. C'est sur la base de la structure de connaissance activée qu'il va, en effet, choisir et organiser l'information qu'il reçoit sur la situation stimulus pour en construire sa propre représentation. Ainsi, l'apprenant va construire une interprétation de la situation problème en fonction de sa propre théorie implicite du pouvoir. Les structures de connaissance ne constituent cependant

pas le seul élément qui entre en ligne de compte dans la construction d'une représentation. D'autres éléments jouent aussi un rôle important, comme l'information disponible sur la situation stimulus et le contexte de production de la représentation (perception du but de la tâche, relation à l'enseignant, etc.), tous ces éléments interagissant étroitement d'ailleurs dans le processus d'élaboration de la représentation. En «aval», l'élaboration de cette représentation peut, le cas échéant, avoir des effets rétroactifs sur la structure de connaissance qui a été activée. Ainsi, l'apprenant va «découvrir» dans cette situation problème un nouveau mode d'exercice du pouvoir dont sa structure de connaissance initiale activée (sa théorie implicite du pouvoir) ne rendait pas compte. Dans ce cas, il pourrait être amené à modifier celle-ci de manière à ce qu'elle puisse rendre compte des éléments nouveaux.

Cette distinction vient donc complexifier quelque peu le modèle général de l'apprentissage présenté plus haut :

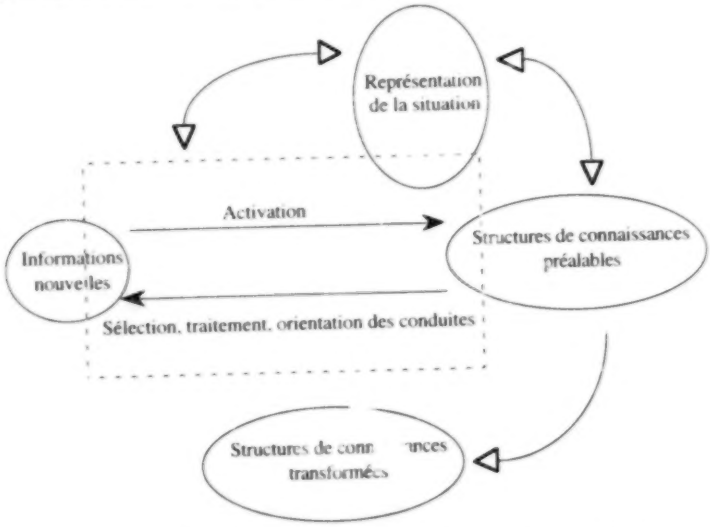


Figure 2 – L'apprentissage dans le paradigme constructiviste (compte tenu de la distinction entre représentation et connaissance)

2. IMPLICATIONS MÉTHODOLOGIQUES DU PARADIGME CONSTRUCTIVISTE

Deux implications méthodologiques fondamentales découlent de ces positions théoriques. Les connaissances ne sont accessibles qu'au travers de représentations circonstancielles et par inférence.

2.1 Accès aux connaissances par les représentations circonstancielles

Tout comme on ne peut agir sur des structures de connaissance (pour les transformer) qu'au travers des représentations circonstancielles dans lesquelles elles sont activées, de même on ne peut accéder à des connaissances (pour les étudier) qu'au travers des représentations dans lesquelles elles sont activées. Les structures de connaissance d'un sujet ne se donnent pas comme telles à l'observation. Elle ne peuvent être qu'inférées inductivement par l'observateur à partir des représentations dans lesquelles elles sont activées par le sujet. Pour reprendre l'exemple donné ci-avant, je ne peux, comme observateur, accéder à la théorie implicite du pouvoir de tel apprenant qu'au travers des représentations circonstancielles qu'il produit et dans lesquelles sa théorie du pouvoir est effectivement activée. Et encore, ce que je pourrai inférer, c'est la théorie du pouvoir que l'apprenant a effectivement activée dans telle ou telle représentation[2].

Dans cette perspective, il est donc essentiel de s'interroger sur la validité et sur la pertinence du dispositif méthodologique de recueil des données mis en œuvre pour étudier des changements au niveau des structures de connaissance : quel champ de représentations – et partant, quel champ de connaissances – ce dispositif est-il susceptible d'activer chez le sujet? Le champ de connaissances effectivement activé chez le

2 C'est tout le problème de la labilité des connaissances qui est ici posé. À cet égard, une des critiques adressées à la distinction entre connaissance et représentation telle qu'elle a été présentée consiste à dire qu'elle ne prend pas en compte l'existence possible de structures de niveau intermédiaire entre, d'une part, des connaissances effectivement stables, peu labiles, stockées en mémoire à long terme, et d'autre part, des représentations purement circonstancielles, produites *hic et nunc* et liées à des situations particulières. Je dois cette judicieuse remarque, en particulier, à Linda Allal.

sujet par le dispositif est-il réellement celui que le chercheur vise à investiguer? Nous reviendrons plus loin sur ce point et sur ses implications concrètes.

2.2 Accès aux connaissances par inférence

Les représentations ne manifestent pas de manière «transparente» les structures de connaissance sur lesquelles elles s'appuient. Rappelons en effet que les représentations ne s'élaborent pas uniquement sur la base de connaissances préalables. D'autres éléments entrent en ligne de compte dans leur élaboration, tels que l'information disponible sur la situation stimulus et le contexte de production de la représentation. De plus, ces éléments interagissent étroitement dans le processus représentationnel. Par exemple, l'apprenant confronté à la même situation-stimulus peut être amené à activer des structures de connaissance différentes – et donc *in fine*, à produire des représentations différentes de cette situation – selon que cela se passe dans le cadre d'un examen, d'une séance de cours ou d'une interview de recherche. De même, le sujet va sélectionner et organiser différemment l'information relative à la situation stimulus selon la structure de connaissance activée. Inversement, il peut être amené à activer des structures de connaissance différentes selon la manière dont la situation stimulus lui est présentée, selon la nature et la quantité de l'information qui lui est effectivement fournie.

Ces considérations ne sont pas sans conséquences importantes, en particulier pour la méthodologie du traitement des données. Elles invitent à la plus grande prudence dans les inférences que l'on peut faire concernant des structures de connaissance à partir de comparaisons synchroniques et diachroniques entre des représentations. En effet, dans quelle mesure les différences observées entre des représentations peuvent être attribuées à des différences au niveau de leurs structures de connaissance – qu'il s'agisse de représentations produites à différents moments par un même sujet confronté à la même situation stimulus ou de représentations produites au même moment par des sujets différents confrontés à la même situation stimulus? Il s'agit, pour le chercheur (et le formateur) d'être particulièrement attentif dans son interprétation aux autres sources possibles de variance, et notamment les différences liées

au contexte dans lequel les représentations ont été produites et celles qui sont liées aux caractéristiques de la situation stimulus elle-même.

Comme annoncé en début de texte, ces considérations ont largement fondé la conception et le choix du dispositif méthodologique de recueil et de traitement des données utilisés dans quelques-uns des travaux réalisés ou en cours dans notre laboratoire. Je voudrais à présent évoquer brièvement ces dispositifs, simplement dans le but d'illustrer concrètement les propos théoriques développés ci-avant. Faut-il le souligner, ces dispositifs ne constituent qu'une réponse possible parmi tant d'autres aux problèmes soulevés plus haut, une réponse contextualisée par les objectifs spécifiques des recherches en question ainsi que des contraintes et ressources du moment.

3. Implications pour le recueil des données

Je voudrais évoquer ici tout particulièrement deux types d'investigation pouvant être envisagés lorsque l'on étudie les processus de transformation de connaissances et de représentations en cours de formation.

3.1 Premier type de recherche

Un premier type de recherche consiste à identifier a) les structures de connaissance propositionnelle spontanément activées par un sujet face à une situation donnée, en dehors du contexte de la formation; b) les changements éventuels intervenus dans ces structures au cours d'une période donnée (par exemple, au cours de la formation ou au-delà) et c) les liens éventuels entre les changements observés et les caractéristiques de l'apprenant et de la situation de formation, en particulier les objets de connaissance traités dans la formation.

Un collègue et moi-même avons réalisé une recherche qui visait précisément ces trois objectifs (Bourgeois et Nizet, 1989). Elle portait sur des étudiants adultes qui suivent un cours de sociologie, dans le cadre d'un programme (pour adultes) de licence en politique économique et sociale à l'Université catholique de Louvain. Ce cours de quarante-cinq heures était organisé en cinq modules, correspondant chacun à un chapitre du cours. Chaque module présentait la même structure chronologique.

Première séance. – Une situation problème est présentée par un étudiant à l'ensemble du groupe. Cette situation est choisie par l'étudiant en concertation avec l'enseignant qui s'assure de la pertinence de la situation par rapport aux apports théoriques qui font l'objet du module. Cette présentation est immédiatement suivie d'un tour de table où chaque étudiant formule «à chaud» son interprétation personnelle de la situation ainsi que ses propositions d'action.

Entre la première et la deuxième séance de cours, les étudiants prennent connaissance du chapitre théorique traité dans le module en question.

Deuxième séance et suivantes. – Les étudiants discutent ces apports théoriques avec l'enseignant et tentent ensuite de les appliquer à la situation problème du premier cours. Ils confrontent enfin les analyses «spontanées» produites par chacun avec celles qui ont été réalisées à l'aide de la théorie enseignée.

Le dispositif de recueil des données pour notre recherche était le suivant.

En un premier temps (T1), les sujets sélectionnés se voyaient présenter lors d'une première interview une situation problème, soit la narration d'un événement présentant une unité de temps, de lieu et d'action. Cette interview eut lieu quelques jours après la première séance du cours en question. La situation proposée lors de l'interview était strictement la même que celle que le sujet avait entendue quelques jours auparavant. Dans l'interview, après avoir rappelé la situation en question, on demandait alors au sujet, comme lors de la première séance de cours, de proposer son interprétation personnelle de l'événement et ses perspectives d'action. L'entretien était conduit de manière très peu directive (questions de clarification et reformulations uniquement).

En un second temps (T2), quelques jours après la dernière séance de cours (soit plus ou moins trois mois plus tard que T1), les mêmes sujets ont été réinterviewés une seconde fois. Cette deuxième interview était structurée strictement de la même manière que la première (T1) : même situation stimulus, mêmes consignes, même type de conduite

d'entretien et même durée d'entretien. Toutefois, les sujets ont été amenés à plusieurs reprises au cours de cette seconde interview à se positionner par rapport à des éléments exprimés par eux lors de la première interview.

À la suite de cette seconde interview, les chercheurs ont analysé le matériau et formulé une série d'hypothèses concernant, d'une part, la nature des changements les plus significatifs observés dans le discours des sujets entre T1 et T2 et, d'autre part, les explications possibles de ces changements.

En un troisième temps (T3), les sujets ont été interviewés une dernière fois, plus ou moins six mois après la seconde interview. Cette troisième interview se déroulait en deux parties. La première s'est déroulée de manière comparable à la seconde (même situation stimulus, mêmes consignes, même type de conduite d'entretien, plus ou moins même durée d'entretien et demande de positionnement par rapport à des éléments exprimés lors de la deuxième interview). Dans la seconde partie, on a soumis à chaque sujet nos hypothèses concernant les changements observés dans son discours entre T1 et T2 et leurs explications possibles. Le matériau produit dans la première partie de l'interview a été traité comme celui des deux premières interviews. Nous avons pu ainsi identifier les changements intervenus dans le discours du sujet à propos de la même situation problème entre T1, T2 et T3. Le matériau produit dans la seconde partie de la troisième interview a été utilisé principalement pour la validation de nos hypothèses.

Par sa grande ouverture, en particulier au niveau des consignes et de la conduite des entretiens, ce type de dispositif a effectivement permis d'identifier (ou plutôt d'inférer) le champ de connaissance spontanément activé chez le sujet pour traiter la situation stimulus. Il nous a dès lors permis de repérer les changements intervenus dans ce champ de connaissances d'un entretien à l'autre et, enfin, d'examiner la relation entre ces changements et certains aspects de la situation de formation, notamment les objets de connaissance qui y ont été traités. On a ainsi pu constater que le champ de connaissances activé par le sujet face à la situation stimulus dans le cadre de l'interview pouvait s'avérer très éloi-

d'entretien et même durée d'entretien. Toutefois, les sujets ont été amenés à plusieurs reprises au cours de cette seconde interview à se positionner par rapport à des éléments exprimés par eux lors de la première interview.

À la suite de cette seconde interview, les chercheurs ont analysé le matériau et formulé une série d'hypothèses concernant, d'une part, la nature des changements les plus significatifs observés dans le discours des sujets entre T1 et T2 et, d'autre part, les explications possibles de ces changements.

En un troisième temps (T3), les sujets ont été interviewés une dernière fois, plus ou moins six mois après la seconde interview. Cette troisième interview se déroulait en deux parties. La première s'est déroulée de manière comparable à la seconde (même situation stimulus, mêmes consignes, même type de conduite d'entretien, plus ou moins même durée d'entretien et demande de positionnement par rapport à des éléments exprimés lors de la deuxième interview). Dans la seconde partie, on a soumis à chaque sujet nos hypothèses concernant les changements observés dans son discours entre T1 et T2 et leurs explications possibles. Le matériau produit dans la première partie de l'interview a été traité comme celui des deux premières interviews. Nous avons pu ainsi identifier les changements intervenus dans le discours du sujet à propos de la même situation problème entre T1, T2 et T3. Le matériau produit dans la seconde partie de la troisième interview a été utilisé principalement pour la validation de nos hypothèses.

Par sa grande ouverture, en particulier au niveau des consignes et de la conduite des entretiens, ce type de dispositif a effectivement permis d'identifier (ou plutôt d'inférer) le champ de connaissance spontanément activé chez le sujet pour traiter la situation stimulus. Il nous a dès lors permis de repérer les changements intervenus dans ce champ de connaissances d'un entretien à l'autre et, enfin, d'examiner la relation entre ces changements et certains aspects de la situation de formation, notamment les objets de connaissance qui y ont été traités. On a ainsi pu constater que le champ de connaissances activé par le sujet face à la situation stimulus dans le cadre de l'interview pouvait s'avérer très éloi-

gné du champ qui était en fait ciblé par l'enseignant dans le cadre du cours. En d'autres termes, les sujets dans la situation d'interview ont fait appel à des connaissances qui s'inscrivent dans un champ plus ou moins éloigné de celui visé par l'enseignant. Cependant, il faut souligner avec insistance que ce constat-là ne permet pas d'inférer que des changements ne se sont pas néanmoins produits chez les sujets dans le champ de connaissances visés par l'enseignant. Par exemple, ce n'est pas parce qu'un sujet n'a pas fait appel à la notion de pouvoir (abordée dans le cadre du cours) pour traiter la situation stimulus que sa conception du pouvoir ne s'est pas transformée au fil de la formation. Simplement, le type de dispositif méthodologique décrit plus haut ne permet pas de le vérifier puisqu'il ne vise pas à activer chez le sujet le champ de connaissances particulier visé par l'enseignant. Il permet cependant de vérifier que des changements se sont (ou non) bien produits dans les structures de connaissance auxquelles le sujet a fait spontanément appel pour traiter la situation.

3.2 Deuxième type de recherche

Le deuxième type de recherche consiste, précisément, à identifier les changements éventuels intervenus chez un sujet dans un champ de connaissances prédéterminé. C'est typiquement le cas d'études qui visent à vérifier si des changements sont intervenus chez les apprenants dans le domaine de connaissances cadré en fonction des objectifs de l'enseignant au cours d'une séquence d'apprentissage et, le cas échéant, d'évaluer ces changements, c'est-à-dire de vérifier si les structures de connaissance qui s'inscrivent dans le champ ciblé se sont transformées dans le sens attendu en fonction des objectifs d'apprentissage.

Une recherche de ce type vient d'être réalisée dans notre laboratoire (Venegas, 1993). L'étude porte sur un cycle de formation organisé par un centre d'éducation populaire au Chili et destiné à des agents de santé communautaire de la banlieue de Santiago. Les objectifs d'apprentissage explicitement définis par les responsables de la formation consistaient, pour faire bref, à développer chez les apprenants une conception plus participative de la politique de santé communautaire sous ses différents aspects (objectifs et stratégies du dispensaire, relations entre les

différents acteurs concernés, etc.). Le but de l'étude consiste, entre autres, à vérifier si les structures de connaissance des sujets dans ce domaine se sont effectivement transformées au cours de la formation et si les transformations se sont opérées dans le sens attendu en fonction des objectifs des responsables de la formation[3].

Le dispositif de recueil de données était le suivant. Une situation problème (narration d'événement) a été préalablement construite par le chercheur, en concertation avec les responsables de la formation. Il s'agissait de construire la situation problème de manière à maximiser la probabilité qu'elle active effectivement chez les sujets le domaine de connaissances visé par la formation (en l'occurrence, leurs conceptions relatives aux différents aspects de la politique de santé communautaire). Pour le reste, le dispositif était très proche de celui décrit plus haut, à une différence (majeure) près : les consignes étaient plus fermées et la conduite de l'entretien plus directive (entretien de type semi-structuré) de façon à susciter au maximum l'émergence des représentations dans le champ ciblé *a priori*. L'exigence majeure à rencontrer dans la conception et la gestion du dispositif de recueil dans ce type de recherche est double. Il s'agit, d'une part, de cadrer suffisamment le champ de représentation de façon à obtenir un maximum d'information sur les connaissances du sujet dans le domaine ciblé. D'autre part, il s'agit en même temps de veiller à ne pas induire dans ce champ des représentations et des structures de connaissances qui ne seraient pas les siennes (mais bien, par exemple, celles du chercheur ou du formateur). Pour reprendre notre exemple, il s'agit donc à la fois d'amener le plus possible le sujet à exprimer ses conceptions de la politique de santé communautaire mais en même temps de ne pas induire une conception particulière, autre que la sienne.

Ce type de dispositif permet donc précisément de faire ce qu'on ne pouvait faire dans le premier, à savoir d'identifier les changements intervenus chez un sujet dans un champ de connaissance prédéterminé. Par contre, l'observation de changements (ou de l'absence de change-

3 En réalité, l'objectif de la recherche allait au-delà de cela puisqu'il s'agissait également de vérifier un certain nombre d'hypothèses concernant des facteurs d'apprentissage dans ce type de situation.

ment) dans le champ de connaissances ciblé *a priori* par le chercheur ne permet en aucun cas de vérifier si des changements – peut-être plus importants encore – se sont ou non produits chez le sujet dans un autre champ de connaissances, tel que celui qu'il aurait pu activer spontanément dans un dispositif de recueil plus ouvert.

4. Implications pour le traitement des données

Je voudrais enfin évoquer la méthodologie de traitement des données qui a été utilisée dans les travaux évoqués plus haut et qui est d'ailleurs toujours en cours d'affinement. Le matériau à traiter est donc du matériau discursif (retranscription intégrale du discours produit par le sujet à propos de la situation stimulus dans le cadre d'une interview). Pour traiter ce type de matériau dans la perspective développée plus haut, il nous fallait nous doter d'un instrument nous permettant d'inférer avec le maximum de rigueur les structures de connaissance propositionnelle qui sous-tendent la représentation exprimée dans le discours du sujet. Nous nous sommes pour cela tournés vers une technique d'analyse du discours, appelée «analyse structurale». Il s'agit d'une technique développée à l'origine par une équipe de sociologues de l'Université catholique de Louvain (Rémy, 1990; Hiernaux, 1977; Ruquoy, 1990; Nizet, 1984), sur la base des travaux de l'école de Greimas (1966) et adaptée par Nizet (1982, 1984)[4]. Il serait hors de propos d'en faire ici une présentation détaillée et je renvoie pour cela le lecteur aux références citées ci-avant[5]. Disons simplement que cette technique permet d'inférer, selon des règles très précises, la structure sémantique profonde du discours dans sa triple dimension informationnelle, structurelle et normative, en tenant compte des conditions de production du discours. Contrairement aux techniques d'analyse thématique, elle s'attache moins à la surface du discours qu'à sa structure sémantique profonde. Cette technique, couplée à ce dispositif de recueil du matériau, s'est révélée particulièrement pertinente pour le type de recherches envisagé ci-dessus, ainsi que

4 La technique fait actuellement l'objet d'une étude systématique de validation qui devrait encore conduire à des améliorations de certains aspects de la technique en vue d'une utilisation pour le type de recherche dont il est ici question.

5 Pour une illustration de l'analyse structurale appliquée à l'étude des changements de représentations, voir Bourgeois (1992a).

nous l'avons argumenté par ailleurs (Bourgeois, 1992a), surtout en comparaison à d'autres instruments disponibles pour l'inférence de représentations.

En effet, d'une part, les techniques quantitatives d'analyse de contenu sont d'une manière générale peu appropriées pour l'étude de représentations individuelles. Qu'il s'agisse des instruments utilisés par les psychologues sociaux qui cherchent à identifier des représentations sociales (c'est-à-dire, par essence, des représentations communes à une population donnée) ou de ceux qu'utilisent les psychologues cognitivistes pour établir des lois générales sur le fonctionnement cognitif, ils ne permettent de réaliser des inférences (par traitement statistique) qu'à partir d'observations sur l'ensemble d'une population. De même, les techniques dites «de rappel», utilisées par les psychologues cognitivistes (pour étudier la mémoire, par exemple), s'avèrent également peu pertinentes ici puisque nous cherchons à étudier des représentations d'un niveau de complexité bien plus élevé que celles qui peuvent être saisies par ces techniques. D'autre part, l'effort méthodologique des didacticiens qui étudient des représentations complexes d'apprenants individuels en situation de formation (essentiellement par des méthodes qualitatives) s'est surtout porté sur ce qu'ils appellent l'évocation des représentations, c'est-à-dire, en fait, la production de matériau représentationnel (discursif, graphique ou comportemental) et bien peu sur le traitement inférentiel de ce matériau (cf., par exemple, Tiberghien, 1983; Osborne et Gilbert, 1980a, 1980b; Jonnaert *et al.*, 1988). Ce travail d'inférence, essentiellement qualitatif, reste donc très largement intuitif. À notre connaissance, peu de ces instruments ont en tout cas fait l'objet d'une étude systématique de validation. De plus, en ce qui concerne le traitement de matériau discursif en particulier, peu d'instruments utilisés par les didacticiens permettent de rendre compte des dimensions structurelle et normative de la représentation.

CONCLUSION

Pour conclure, je voudrais livrer brièvement une réflexion qui a fait l'objet d'une discussion particulièrement riche au sein du groupe des auteurs de cet ouvrage. Le chercheur ou l'enseignant qui se penche

sur les représentations et sur les connaissances de l'apprenant est confronté au paradoxe suivant.

D'une part, sur le plan épistémologique, il doit faire preuve d'une grande lucidité concernant le statut de son objet d'étude. Il devra admettre en dernière instance qu'il ne pourra jamais que se représenter, c'est-à-dire construire mentalement les structures de connaissance et de représentation de l'apprenant. Celles-ci ne constituent pas des objets réels, inscrits dans la réalité mentale du sujet. Elles ne sont jamais que le résultat d'un processus de construction mentale qui s'élabore dans un contexte donné, selon les règles et procédures méthodologiques que l'observateur s'est donné et qui lui permettront de reconstruire le plus fidèlement possible le sens du discours (ou, le cas échéant du comportement) observé. L'observateur ne «découvre» pas les connaissances ni les représentations du sujet. Il se les représente, il les infère. Étudier des représentations et des connaissances, c'est avant tout produire du sens. Il devra par conséquent veiller en permanence à éviter le piège de la réification de son objet. Dans le même ordre d'idée, le concept de connaissance, comme structure cognitive stabilisée en mémoire à long terme, est avant tout une construction théorique. Quelle réalité recouvre d'ailleurs le concept de mémoire à long terme? Comme cela a été dit plus haut, les connaissances ne sont pas des objets mentaux, qui se donneraient comme telles à l'observation. Elles ne sont accessibles que par inférence, au travers des représentations dans lesquelles elles sont effectuées. De même, la distinction entre des connaissances ainsi définies et des représentations qui seraient par définition des structures circonstancielles, extrêmement labiles est également une construction théorique. Comme cela a été remarqué plus haut, il serait sans doute plus correct de considérer ces deux concepts comme des balises qui aident à penser et à agir comme des pôles théoriques entre lesquels on peut concevoir une multitude de structures intermédiaires, *plus ou moins* labiles, *plus ou moins* liées à situations particulières.

D'autre part, lorsqu'il est soumis aux contraintes propres à l'action – que ce soit l'action d'étudier les structures de connaissances ou l'action d'enseigner, c'est-à-dire de transformer ces connaissances – le chercheur ou le formateur est tenté, sinon forcé, d'«oublier», au moins

provisoirement, le statut théorique de ces concepts balises, de les réifier, de leur conférer un statut ontologique de réalité, de choses. Il est souvent amené à faire «comme si» les structures de connaissance ou de représentation qu'il a identifiées existaient «réellement» dans la tête du sujet, indépendamment des conditions méthodologiques et du contexte dans lesquels elles ont été identifiées. Il est souvent contraint de faire «comme si» les connaissances pouvaient exister réellement en dehors des représentations dans lesquelles elles sont mises en œuvre, «comme si» les représentations existaient réellement en dehors du discours du sujet, etc. En d'autres termes, le chercheur ou le formateur qui se penche sur les connaissances et sur les représentations de l'apprenant est quelque sorte condamné à la fois à une lucidité épistémologique et à une certaine naïveté pragmatique pour traiter son objet.

RÉFÉRENCES

BOURGEOIS, É. (1992a).
L'analyse structurale comme instrument de description et d'évaluation des changements de représentations des apprenants. Communication au colloque de GRIFE sur le thème *Représentations, images mentales et modèles*. Montpellier, 13-14-15 janvier 1992. Actes du colloque (sous presse).

BOURGEOIS, É. (1992b).
Connaissances, représentations et discours : un cadre conceptuel pour l'évaluation des effets de formation. Communication au XXV^e^ Congrès international de psychologie. Bruxelles, juillet 1992. Version provisoire publiée dans les cahiers *Pédagogies* du Laboratoire de pédagogie expérimentale de l'Université catholique de Louvain, Vol. 3, juillet, 63-78.

BOURGEOIS, É. et NIZET, J. (1989).
Formation expérientielle, changements de représentations et trajectoire sociale : Une étude de cas. Communication au IV^e^ Symposium du RIFREP. Paris, 19-22 avril 1989. Actes du symposium.

CORDIER, F., DENHIÈRE, G., GEORGE, C., CRÉPAULT, J., HOC, J.M. et RICHARD, J.F. (1990).
Connaissances et représentations. *In* Richard, J.F., Bonnet, C. et Ghiglione, R. (éd.), *Traité de psychologie cognitive, tome 2. Le traitement de l'information symbolique* (p. 33-102). Paris : Dunod.

GREIMAS, A.J. (1966).
Sémantique structurale. Paris : Larousse.

HIÉRNAUX, J.P. (1977).
L'institution culturelle II. Méthode de description structurale. Louvain : Presses universitaires de Louvain.

JONNAERT, P., DUQUESNE, F. et TOURNEUR, Y. (1988).
Vers une typologie des techniques d'émergence des préacquis cognitifs. *Mesure et évaluation en éducation, 12*(1), 41-69.

NIZET, J. (1982).
Initiation à l'analyse structurale du discours. Namur : Facultés universitaires Notre-Dame de la Paix, Cahiers de la Faculté des sciences économiques et sociales, *4*, janvier.

NIZET, J. (1984).
Manuel d'analyse structurale du discours. Namur : Facultés universitaires Notre-Dame de la Paix. Document de travail non publié.

OSBORNE, R.J. et GILBERT, J.K. (1980a).
A technique for exploring students' views of the world. *Physics Education, 15*, 376-379.

OSBORNE, R.J. et GILBERT, J.K. (1980b).
A method for investigating concept understanding in science. *European Journal of Science Education, 2*(3), 311-321.

RÉMY, J. (1990).
L'analyse structurale et la symbolique sociale. *In* Rémy, J. et Ruquoy, D. (éd.), *Méthodes d'analyse de contenu et sociologie* (p. 111-132). Bruxelles : Publications des Facultés universitaires Saint-Louis.

RICHARD, J.F. (1990).
Les activités mentales : comprendre, raisonner, trouver des solutions. Paris : Armand Colin.

RUQUOY, D. (1990).
Les principes et procédés méthodologiques de l'analyse structurale. *In* Rémy, J. et Ruquoy, D. (éd.), *Méthodes d'analyse de contenu et sociologie* (p. 93-109). Bruxelles : Publications des Facultés universitaires Saint-Louis.

TIBERGHIEN, A. (1983).
Revue critique sur les recherches visant à élucider le sens des notions de circuits électriques pour les élèves de 8 à 20 ans. Research on Physics Education (Proceedings of the First International Workshop), 91-107. Paris : Éditions du CNRS.

VENEGAS, P. (1993).
Conflits sociocognitifs et changement de représentation en formation d'adultes : une étude de cas. Thèse de doctorat, Université catholique de Louvain.

CROYANCES SAILLANTES ET CONDITIONS DE PRATICABILITÉ DE L'ÉVALUATION FORMATIVE DES APPRENTISSAGES TELLES QUE PERÇUES PAR LES ENSEIGNANTS

SERGE P. SÉGUIN
CARMEN PARENT
ROSE BURELLE
Université du Québec à Montréal

INTRODUCTION

L'étude rapportée ici a pour but de contribuer à comprendre les raisons qui motivent les enseignants à adopter ou non diverses pratiques d'évaluation formative dans leurs classes. La partie étudiée de ces déterminants concerne les croyances et les attitudes des enseignants à l'égard de pratiques souhaitées en ce domaine. L'instrument utilisé a été construit afin de mesurer les composantes du modèle emprunté à la théorie de l'action raisonnée (Fishbein et Ajzen, 1975; Ajzen et Fishbein, 1980) : ce modèle a été retenu pour expliquer ou comprendre le comportement «pratiquer l'évaluation formative» chez les enseignants des ordres d'enseignement primaire et secondaire.

Avant de décrire les échantillons de répondants, de présenter les résultats et de les discuter en regard du modèle retenu et du but de l'étude, il convient de rappeler brièvement les composantes du modèle théorique et de l'instrument, puis de définir le concept de croyances saillantes (*salient beliefs*).

1. RAPPEL DU MODÈLE

Burelle, Gadbois, Parent et Séguin (1991) ont déjà examiné et exposé les composantes du modèle retenu, concluant que ce modèle semblait offrir une voie utile pour expliquer des comportements humains susceptibles d'être modifiés.

À condition que l'intention, l'attitude et la croyance soient toutes exprimées en termes du comportement à l'étude et de ses conséquences,

et partant du postulat que l'intention de manifester le comportement est le seul prédicteur de ce comportement, le modèle retenu a été proposé pour expliquer ou comprendre : a) l'intention par des attitudes, ces attitudes étant pondérées par une norme subjective définie par un degré de motivation à se conformer à des croyances normatives ou à des attentes de référents telles qu'elles sont perçues; b) les attitudes par des croyances relatives aux conséquences de manifester le comportement et pondérées par des évaluations subjectives de ces conséquences. On augmente ainsi la compréhension de la chaîne qui relie les croyances au comportement, mais sans augmenter la prédiction du comportement. Pour modifier un comportement, il faut intervenir sur l'ensemble des composantes et de leurs relations de façon à assurer une intention stable et congruente à la manifestation du comportement souhaité, soit, ici, la pratique de l'évaluation formative des apprentissages des élèves en classe.

2. Rappel de l'instrument

L'instrument utilisé est celui qui a été construit, validé et décrit par Gadbois, Burelle, Parent et Séguin (1991). Après une étude conceptuelle élaborée et retenant la définition de Scallon (1988, 155), puis après une étude empirique de confrontation avec la praxis d'enseignants et de conseillers pédagogiques, ces auteurs ont retenu 50 conséquences qui définissent le domaine du comportement «pratiquer l'évaluation formative». Quatre de ces conséquences ne sont pas retenues ici dont les deux qui sont susceptibles de ne s'appliquer qu'au secondaire. Les deux autres conséquences non conservées sont celles que les enseignants des deux ordres d'enseignement perçoivent comme n'appartenant pas au domaine de l'évaluation formative, à savoir que cette pratique 1) les empêcherait de couvrir tous les objectifs des programmes et 2) valoriserait la profession enseignante dans la société. Les termes *praticabilité* et *applicabilité* utilisés dans le texte sont jugés comme synonymes. Cependant, pour éviter la confusion et pour respecter la terminologie utilisée dans les travaux précédents, on emploiera le terme *praticabilité* lorsqu'il s'agit d'évaluation formative dans sa globalité et *applicabilité* lorsqu'il est question des conséquences de la pratique de l'évaluation formative.

Les 46 conséquences retenues pour l'étude sont regroupées en cinq sous-domaines ou niveaux d'impact, tel qu'ils apparaissent en colonne de gauche du tableau 5 (cf. *infra*).

Dans le livret du questionnaire, le répondant est invité à fournir trois réponses à l'égard de chacune des conséquences identifiées, comme suit.

Dans une première partie, il doit indiquer son degré d'accord sur une échelle en 7 points (-3 à 3) quant à l'appartenance de la conséquence au domaine de l'évaluation formative; [c'est l'échelle dite d'appartenance des conséquences, permettant d'identifier les plus (et les moins) saillantes, c'est-à-dire celles qui, selon les répondants, appartiennent le plus ou le moins au comportement étudié; comme les deux conséquences dont l'appartenance a été cotée négativement ont été ignorées, l'échelle d'appartenance utilisée dans l'étude va de 9 à 3.

Dans une deuxième partie, il doit indiquer son évaluation des désavantages et des avantages de la conséquence; l'échelle des réponses qui est en 7 points va de -3 à 3; toutefois, une seule conséquence a été cotée négativement (désavantageuse) par les enseignements des deux ordres d'enseignement : il s'agit de l'augmentation de la tâche générale de l'enseignant qui résulte de la pratique de l'évaluation formative. Il doit enfin indiquer, sur une échelle en 5 points (0 à 4), son appréciation du degré d'applicabilité de la conséquence dans le contexte scolaire actuel.

Cette composante d'applicabilité n'est toutefois pas présente dans le modèle d'Ajzen et Fishbein. Elle a été introduite aux fins de l'étude rapportée ici.

Le ministère de l'Éducation du Québec, les conseillers pédagogiques et la direction de l'école sont les trois référents retenus pour mesurer les croyances normatives des répondants ainsi que leur degré subjectif de motivation à s'y conformer, chacune de ces deux variables étant exprimée sur une échelle en 7 points (-3 à 3). Quant à l'attitude proprement dite, elle est mesurée par trois échelles, chacune en 7 points (-3 à 3), en termes de décourageant-stimulant, absurde-sage, nuisible-bénéfique.

3. CROYANCES SAILLANTES

Dans la théorie de l'action raisonnée, le mot croyance signifie globalement l'information que détient un sujet sur un objet donné, ici le comportement «pratique l'évaluation formative». Selon Ajzen et Fishbein (1980), même si une personne possède de nombreuses croyances à propos d'un objet donné, le nombre de croyances présentes dans son champ de conscience à un moment donné est assez restreint, soit entre 5 et 9 selon leur estimation. La théorie de l'action raisonnée établit que ces croyances dites saillantes sont les déterminants immédiats de l'attitude et qu'il n'est pas très utile d'amener un sujet à prendre en compte un nombre plus important de ses croyances. En pratique, il suffit donc au chercheur désireux de connaître l'attitude d'une personne envers un objet donné de mettre au point une procédure qui lui permet de repérer au sein d'un ensemble de croyances lesquelles sont saillantes et lesquelles peuvent être négligées. Il n'existe aucune règle fixe pour tracer une ligne de démarcation entre les croyances saillantes et les autres, l'estimation de 5 à 9 représentant un ordre de grandeur.

Donc, lorsqu'il s'agit d'identifier les croyances saillantes d'un groupe, il est admissible de retenir celles auxquelles sont associées les cotes moyennes les plus élevées. L'instrument utilisé ici permet d'obtenir des résultats sur trois échelles relativement aux croyances comportementales : appartenance au domaine de l'évaluation formative, évaluation des implications de sa pratique, applicabilité dans le contexte scolaire actuel.

Parmi les 46 conséquences, celles qui reçoivent les plus hautes cotes des répondants sur l'échelle d'appartenance sont celles qu'ils identifient comme les effets les plus plausibles de l'évaluation formative : ce sont les conséquences jugées les plus pertinentes ou importantes de pratiquer l'évaluation formative. Les croyances saillantes sont donc ici définies comme les croyances comportementales relatives aux conséquences qui reçoivent les cotes les plus élevées sur l'échelle d'appartenance. D'autres façons de définir les croyances saillantes seront examinées et comparées dans une autre étude de Parent, Séguin, Burelle (en préparation).

Pour comprendre la formation ou l'expression d'une attitude à l'égard de l'évaluation formative, les échelles d'évaluation et d'applicabilité ne sont alors d'intérêt que pour les conséquences considérées comme saillantes. Il importe donc d'étudier les valeurs, sur les échelles d'évaluation et d'applicabilité, des conséquences qui définissent les croyances comportementales saillantes : ces dernières ne sont pas nécessairement les plus positivement évaluées ni jugées les plus applicables.

4. ÉCHANTILLONS

L'instrument de mesure a été administré au printemps 1990. Le bassin cible de l'étude correspond au Montréal métropolitain, à savoir les régions administratives Nord de Montréal (6.1), Sud de Montréal (6.2) et Île-de-Montréal (6.3). Pour chacun des ordres d'enseignement (primaire et secondaire), la carte démographique de la population-cible a été préparée à partir des statistiques les plus récentes du ministère de l'Éducation et les strates de la population ont été déterminées en tenant compte des variables suivantes : sexe, niveau de scolarité et nombre d'années d'expériences. Deux échantillons, un pour le primaire et l'autre pour le secondaire, chacun de grandeur 200, ont ensuite été tirés de façon à ce que les sujets soient proportionnellement répartis dans chacune des cinquante commissions scolaires du bassin cible. Le contact avec les enseignants des échantillons s'est effectué par l'intermédiaire des conseillers en mesure et évaluation de chacune des commissions scolaires, sauf pour la Commission des écoles catholiques de Montréal dont les enseignants ont été rejoints directement par un des chercheurs de l'équipe. Les conseillers ont reçu les profils socioprofessionnels des enseignants échantillonnés dans leur commission scolaire et ils ont procédé aléatoirement au choix des sujets conformes à ces profils. Les conseillers ont également veillé à la distribution des questionnaires qui ont été retournés à l'équipe de recherche, soit directement, soit par l'intermédiaire des conseillers. Au secondaire, 58,5 % (117/200), et au primaire, 64,5 % (129/200), des enseignants ainsi sollicités ont retourné leur questionnaire complété. La représentativité de cet échantillon réel a été vérifiée et trouvée suffisante au regard du plan d'échantillonnage.

Le tableau 1 présente la répartition des répondants selon les variables de stratification. On note une très grande majorité féminine au

primaire (5 contre 1) et une majorité masculine moindre au secondaire (2 contre 2). Les répondants ont, en moyenne, 44 ans d'âge au primaire et 46 ans au secondaire; 22 ans d'expérience dans l'enseignement au primaire et 23 ans au secondaire; 16 ans de scolarité au primaire et 17 ans au secondaire.

Tableau 1 – Répartition des échantillons des répondants selon le sexe, l'expérience de travail et le niveau de scolarité

Caractéristique	Primaire		Secondaire	
	f	%	*f*	%
SEXE				
femmes	98	83,8	34	34,3
hommes	19	16,2	65	65,7
(inconnu)	(12)	-	(18)	-
EXPÉRIENCE				
1 à 9 ans	5	4,2	4	3,8
10 à 19 ans	40	33,9	27	25,7
20 à 29 ans	54	45,8	54	51,4
30 ans et plus	19	16,1	20	19,1
(inconnue)	(11)	-	(12)	-
SCOLARITÉ				
14 ans ou moins	21	17,6	8	7,7
15 ans	16	13,4	6	5,7
16 ans	23	19,3	20	19,2
17 ans	27	22,7	19	18,3
18 ans	19	16,0	19	18,3
19 ans	12	10,1	30	28,8
20 ans	1	0,8	1	1,0
21 ans	0	0	1	1,0
(inconnue)	(10)	-	(13)	-
Nombre de répondants :	Np = 129		Ns = 117	

f indique la fréquence brute.
% indique la fréquence relative en excluant les sujets dont la caractéristique est inconnue.

5. RÉSULTATS

Les résultats présentés et discutés dans cette section sont disposés selon les ordres d'enseignement placés en parallèle, dans l'ordre suivant : l'intention, l'attitude, les croyances normatives et la motivation à s'y conformer, puis les croyances comportementales.

5.1 L'intention

Les enseignants ont une intention positive, mais mitigée au primaire (moyenne Mp = 1,5) et plutôt faible au secondaire (Ms = 1,0), de pratiquer l'évaluation formative des apprentissages de leurs élèves (voir tableau 2). Cette intention varie beaucoup d'un enseignant à l'autre, comme l'indique la grandeur des écarts types (Sp = 1,6 et Ss = 1,9) et les distributions des réponses.

Tableau 2 – Intention comportementale*

Énoncé	Échelle	Primaire % et stat.	Secondaire % et stat.
Je suis prêt à pratiquer l'évaluation formative des apprentissages de mes élèves.	-3	5,5	7,9
	-2	4,7	10,5
	-1	0,8	2,6
	0	4,7	13,2
	+1	15,6	9,6
	+2	43,8	31,6149
	+3	25,0	24,6
	N	128	114
	M	1,51	0,99
	S	1,60	1,94

* Note relative aux tableaux 2 à 7. Sauf pour l'échelle d'applicabilité qui va de 0 à 4 (dernière colonne des tableaux 5 à 7), toutes les autres échelles sont en sept points (de -3 à 3) comme suit.

Valeur intensité		Pôles indiqués selon chaque énoncé			
-3	fortement	désaccord	décourageant	absurbe	nuisible
-2	moyennement	désaccord	décourageant	absurbe	nuisible
-1	légèrement	désaccord	décourageant	absurbe	nuisible
0	neutre ou indécis	-	-	-	-
1	légèrement	accord	stimulant	sage	bénéfique
2	moyennement	accord	stimulant	sage	bénéfique
3	fortement	accord	stimulant	sage	bénéfique

En outre, la légende suivante s'applique pour ces tableaux :

stat. % : fréquence relative des réponses exprimées;
N : nombre des réponses exprimées (excluant les omissions);
M : valeur moyenne observée des réponses exprimées;
S : écart type de la distribution réponses exprimées.

L'examen des pourcentages indique que 84,4 % des enseignants du primaire et 65,8 % des enseignants du secondaire se disent prêts, à des degrés divers, à pratiquer l'évaluation formative. Toutefois, en considérant les valeurs extrêmes +2 et +3 de l'échelle, environ les deux tiers des enseignants du primaire et la moitié de deux du secondaire seraient prêts à pratiquer l'évaluation formative, on constate d'une manière plutôt systématique, alors qu'environ le dixième des enseignants du primaire et le sixième de ceux du secondaire (valeurs -3 et -2 de l'échelle) n'auraient pas du tout l'intention de le faire.

L'intention de pratiquer l'évaluation formative est donc systématiquement plus faible au secondaire qu'au primaire, sauf pur l'intention positive maximum (valeur 3) qui anime le quart des enseignants à chacun des deux ordres d'enseignement.

Comment expliquer ou comprendre, d'une part, cette intention positive pour certains et, d'autre part, l'intention négative ou absente du sixième des enseignants du primaire et du tiers de ceux du secondaire?

5.2 L'attitude

L'attitude est mesurée par trois échelles, chacune définie par deux pôles extrêmes. Les résultats (voir tableau 3) semblent démontrer que ces trois échelles mesureraient des aspects différents de l'attitude.

L'échelle «absurde-sage» (b) conduit à des résultats dont les distributions, caractéristiques et interprétations sont très semblables à celles des résultats sur la variable intention. Comme les termes utilisés pour décrire cette échelle sont les plus généraux, ils laissent les répondants enseignants exprimer leur attitude en fonction de conséquences générales d'ensemble tout comme les invitait à le faire l'énoncé utilisé pour l'expression de l'intention.

Les autres échelles font appel à des sous-domaines plus spécifiques des conséquences. Ainsi, l'échelle «décourageant-stimulant» (a) semble faire davantage appel aux conséquences qui auraient un impact sur leur perception des exigences de leurs tâches qui en découleraient et sur leur perception des énergies professionnelles qu'il leur faudrait alors

Tableau 3 – Attitude envers le comportement

Énoncé	Échelle	Primaire	Secondaire
Pratiquer l'évaluation formative dans le contexte scolaire actuel est		% et stat.	% et stat.
a) décourageant-stimulant	-3	4,9	10,7
	-2	8,2	15,2
	-1	14,8	8,9
	0	1,6	7,1
	+1	11,5	13,4
	+2	43,4	27,7
	+3	15,6	17,0
	N	122	112
	M	0,99	0,48
	S	1,79	2,06
b) absurde-sage	-3	3,5	7,8
	-2	1,8	6,9
	-1	9,7	8,8
	0	11,5	10,8
	+1	13,3	12,7
	+2	31,9	27,5
	+3	28,3	25,5
	N	113	102
	M	1,38	0,95
	S	1,60	1,92
c) nuisible-bénéfique	-3	0,9	4,0
	-2	0,0	2,0
	-1	0,9	2,0
	0	10,6	15,8
	+1	20,4	14,9
	+2	27,4	34,7
	+3	39,8	26,7
	N	129	101
	M	1,91	1,47
	S	1,15	1,51

investir. Cette explication est plausible lorsqu'on compare les résultats sur cette échelle (a) d'attitude aux résultats sur l'échelle (b) d'attitude ou sur l'échelle d'intention. En effet, pour chacun des ordres d'enseignement, on note une baisse de la moyenne d'environ 0,5, une augmentation d'écart type d'environ 0,2, une baisse d'environ 10 % en points de pourcentage à la valeur 3, accompagnée d'une hausse d'environ 15 % en points de pourcentage aux valeurs négatives. Les enseignants trouvent donc le défi moins stimulant, ou plus décourageant, qu'ils se disent prêts à relever le défi : le manque de stimulation professionnelle, ou l'impression d'un certain découragement devant les tâches à accomplir, apparaît donc comme un motif modérateur ou suppressif de vouloir pratiquer l'évaluation formative, sauf pour le sixième des enseignants de chaque ordre qui trouvent le défi fortement stimulant.

Quant à l'échelle «nuisible-bénéfique» (c), elle semble appeler l'attention des répondants sur les conséquences du comportement en termes de bénéfices qui en résulteraient pour l'élève (voir tableau 7 et discussions plus loin sur les croyances saillantes). C'est sur cette échelle (c) que l'expression de l'attitude des enseignants est la plus élevée, pour chacun des deux ordres d'enseignement : environ 88 % des enseignants du primaire et 76 % de ceux du secondaire expriment alors une attitude positive (≥ 1); 67 % et 51 %, respectivement, trouvent que, dans le contexte scolaire actuel, pratiquer l'évaluation formative est moyennement ou fortement bénéfique (≥ 2).

En résumé, à propos de l'attitude, sur chacune des trois échelles, comme pour l'intention, la moyenne des enseignants du primaire est d'environ 0,5 supérieure à celle des enseignants du secondaire; de plus, c'est sur l'échelle «décourageant-stimulant» que l'attitude exprimée est le moins favorable (la moyenne au primaire correspond à stimulant) : le découragement devant les tâches à accomplir semble donc susceptible de freiner l'intention de pratiquer l'évaluation formative; enfin, c'est sur l'échelle «nuisible-bénéfique» que l'attitude exprimée est le plus favorable (les moyennes observées sont d'une catégorie ou d'une unité plus grandes que les moyennes observées sur l'échelle «décourageant-stimulant») : l'attitude des enseignants apparaît d'autant favorable à l'évaluation formative qu'elle s'appuie sur des critères de bénéfices pour les élèves.

5.3 Les croyances normatives

Les croyances normatives (CN) sont exprimées en termes d'attentes de référents telles qu'elles sont perçues par les répondants (voir tableau 4). Le terme «référents» désigne des personnes ou des groupes qui, pour une raison ou une autre, sont importants aux yeux du sujet et sont perçus comme exerçant sur lui, implicitement ou explicitement, des pressions qui l'incitent à faire une action ou à s'en abstenir. Trois référents sont retenus : le ministère de l'Éducation, les conseillers pédagogiques (du répondant) et la direction de l'école (du répondant).

L'examen des colonnes (CN) du tableau 4 indique que les enseignants du primaire perçoivent au même niveau les attentes des divers référents : la moitié ou plus d'entre eux perçoivent ces attentes comme fortes (valeur 3) et près de 90 % les perçoivent de moyennes à fortes (valeur 2 et 3). Ils font voir ensuite que les enseignants du secondaire perçoivent de la même façon (cote moyenne de 1 mais 50 % les perçoivent de moyennes à fortes) les attentes du ministère de l'Éducation et des directions d'écoles; toutefois, ils perçoivent les attentes de leurs conseillers pédagogiques comme plus élevées (cote moyenne de 1,4 et 60 % qui les perçoivent de moyenne à fortes).

5.4 La motivation

La perception des attentes des référents professionnels n'a d'importance que si elle est accompagnée ou pondérée par un degré de motivation (MOT) à se conformer à ces attentes telles qu'elles sont perçues.

L'examen des colonnes MOT du tableau 4 indique d'une part que les enseignants du primaire expriment un haut degré de motivation (cote moyenne d'environ 2 et plus de 70 % de moyennement à fortement) à se conformer aux attentes, particulièrement aux attentes des directions d'écoles (85 % de moyennement à fortement) et, d'autre part, que les enseignants du secondaire expriment beaucoup plus de réticence à se conformer aux attentes des référents : cote moyenne de 1,6 et 68 % de moyennement à fortement aux attentes des directions d'écoles; cote moyenne de 1,3 et 60 % de moyennement à fortement aux attentes du ministère de l'Éducation; cote moyenne de 1,2 et à peine plus de 55 % de moyennement à fortement aux attentes des conseillers pédagogiques.

Tableau 4 – Croyances normatives (CN) et motivation (MOT) à se conformer aux attentes des référents

Référent*		Primaire		Secondaire	
		CN	MOT	CN	MOT
		% et stat.	% et stat.	% et stat.	% et stat.
a) le ministère de l'Éducation	-3	1,6	0,0	8,0	6,0
	-2	4,8	0.0	9,7	3,4
	-1	5,6	3,1	2,7	5,2
	0	4,8	10,9	9,7	7,8
	+1	5,6	14,1	15,9	17,2
	+2	21,8	36,1	21,2	33,6
	+3	55,6	25,8	32,7	26,7
	N	124	128	113	116
	M	1,96	1,81	1,11	1,35
	S	1,59	1,04	1,98	1,70
b) les conseillers de ma commission scolaire	-3	3,2	0,8	3,6	6,1
	-2	5,6	0,8	7,1	5,2
	-1	3,2	3,1	2,7	7,0
	0	6,4	7,1	13,4	8,7
	+1	2,4	11,0	12,5	16,5
	+2	17,6	46,5	25,9	33,0
	+3	61,6	30,7	34,8	23,5
	N	125	127	112	115
	M	1,98	1,890	1,41	1,17
	S	1,70	1,14	1,72	1,75
c) la direction de mon école	-3	5,6	0,8	7,1	4,3
	-2	2,4	0,0	4,5	0,9
	-1	1,6	3,1	5,4	6,1
	0	4,8	7,0	15,2	7,8
	+1	6,4	3,9	17,0	12,1
	+2	29,6	50,2	23,2	37,4
	+3	49,6	35,2	27,7	31,3
	N	125	128	112	115
	M	1,91	2,03	1,11	1,60
	S	1,64	1,08	1,81	1,54

* - Les énoncés de croyances normatives :
«D'après moi, (le référent préconise la pratique de/pense(nt) que je devrais pratiquer) l'évaluation formative des apprentissages».
- Les énoncés de motivation à se conformer aux attentes du référent :
«Dans le cadre de l'évaluation des apprentissages, je me conforme à ce que (le référent) attend(ent) de moi».
- Toutes les échelles de réponses sont en termes de «désaccord-accord».

En résumé, quant aux attentes perçues et à la motivation à s'y conformer, les enseignants du primaire démontrent une constance élevée (cote moyenne d'environ 2) sur chacune des deux variables et pour chacun des trois référents, mais ils affichent avec une légère préférence à se conformer aux attentes des directions d'écoles.

Cette dernière préférence est le seul point que partagent les enseignants du secondaire avec ceux du primaire. Par ailleurs, les attentes des directions d'écoles secondaires quant à la pratique de l'évaluation formative sont, soit plus faibles que celles des directions d'écoles primaires, soit moins communiquées à leurs enseignants qui le perçoivent comme moins élevées. L'élément majeur relatif aux enseignants du secondaire concerne le paradoxe de leurs relations avec leurs conseillers pédagogiques : selon les enseignants, ce sont ces mêmes conseillers qui préconisent le plus la pratique de l'évaluation formative, mais c'est à leurs attentes que les enseignants sont le moins disposés à se conformer.

5.5 Les croyances comportementales

Pour mesurer les croyances comportementales, les 46 énoncés retenus expriment autant de conséquences de pratiquer l'évaluation formative sur les aspects suivants.

- les apprentissages de l'élève (12 énoncés) : par exemple, cela «aide l'élève à faire le lien entre les divers éléments d'apprentissage acquis»;
- l'organisation de la classe (7 énoncés) : par exemple, cela «entraîne le respect du rythme d'apprentissage de chacun des élèves»;
- le partage des tâches entre collègues (3 énoncés) : par exemple, cela «oblige à diminuer le nombre d'élèves par groupe»;
- le statut de l'enseignant (3 énoncés) : par exemple, cela «conduit l'enseignant à assumer la responsabilité de l'évaluation des apprentissages de ses élèves»;
- les interventions de l'enseignant (21 énoncés) : par exemple, cela «oblige l'enseignant à observer de façon continue ses élèves en cours d'apprentissage» ou encore cela «amène l'enseignant à intervenir de façon continue pour améliorer l'apprentissage en cours».

Les conséquences relatives à ce dernier niveau d'impact peuvent n'énoncer qu'une intervention de l'enseignant, comme dans l'avant-dernier exemple, ou comprendre aussi une indication de l'effet de l'intervention sur les apprentissages des élèves, comme dans l'exemple dernier exemple.

Les énoncés cotés négativement sur l'échelle d'appartenance ayant été exclus de l'étude, les valeurs observées avec cette échelle varient de 0,3 à 2,5 avec une moyenne de 1,6 au primaire et de 0,1 à 2,6 avec une moyenne aussi de 1,6 au secondaire. Sur l'échelle d'évaluation des avantages et désavantages des 46 conséquences, un seul énoncé est coté négativement (de moyennement à fortement désavantageux), à savoir les valeurs moyennes suivantes pour cette conséquence :

Conséquence : ENTRAÎNE UNE AUGMENTATION DE LA TÂCHE

Ordre	Appartenance	Évaluation	Applicabilité
primaire	1,91	-2,25	0,89
secondaire	2,17	-2,48	1,31

Les enseignants jugent que pratiquer l'évaluation formative a pour conséquence probable (moyenne de 2 sur une échelle de 0 à 3) d'augmenter leur tâche, ce qu'ils jugent par ailleurs très désavantageux et perçoivent à peine applicable dans le contexte scolaire actuel.

Les perceptions conflictuelles ou paradoxales relatives à cette conséquence ont sans doute une influence importante sur l'attitude et sur l'intention des enseignants à l'égard de leurs pratiques d'évaluation formative. Si on exclut cette conséquence, les valeurs observées sur l'échelle d'évaluation varient de 0,6 à 3,0 avec une moyenne de 2,2 au primaire et de 0,9 à 2,9 avec une moyenne aussi de 2,2 au secondaire.

Sur l'échelle d'applicabilité (0 = pas du tout; 1 = à peine; 2 = moyennement; 3 = largement; 4 = entièrement), les valeurs observées varient de 0,9 à 2,7 avec une moyenne de 1,9 au primaire et de 0,8 à 2,3 avec une moyenne de 1,6 au secondaire. Les valeurs moyennes à chacune des trois échelles pour chacun des deux ordres d'enseignement sont inscrites au tableau 5, pour l'ensemble et par niveau d'impact.

Tableau 5 – Moyennes observées (M) sur les échelles relatives aux (N) conséquences comportementales regroupées par niveaux d'impact

Niveau d'impact	(N)	Ordre d'ens.	Appartenance (0 à 3)	Évaluation (-3 à 3)	Applicabilité (0 à 4)
Apprentissages de l'élève	(12)	pr.	1,28	2,29	2,01
		sec.	1,21	2,16	1,71
Organisation de la classe	(7)	pr.	1,21	2,32	1,88
		sec.	1,16	2,25	1,54
Partage des tâches entre collègues	(3)	pr.	1,41	1,02	0,92
		sec.	1,71	1,06	1,06
Statut de l'enseignant	(3)	pr.	1,63	2,47	1,85
		sec.	1,63	2,14	1,68
Interventions de l'enseignant	(21)	pr.	1,92	2,22	1,91
		sec.	1,97	2,27	1,67
Total	(46)	pr.	1,59	2,19	1,86
		sec.	1,61	2,17	1,62

L'examen de ce tableau appelle les constatations suivantes. Premièrement, les enseignants des deux ordres sont semblables entre eux sur les échelles d'appartenance et d'évaluation, alors que sur l'échelle d'applicabilité et pour quatre des cinq niveaux d'impact, les enseignants du secondaire trouvent les conséquences de 0,2 à 0,3 moins applicables que ne le trouvent ceux du primaire. Deuxièmement, pour chacun des deux ordres, les énoncés les plus conséquents (échelle d'appartenance) portent sur les interventions de l'enseignant (voir les croyances saillantes en section suivante), alors que les moins conséquents portent sur l'organisation de la classe. Troisièmement, pour chacun des deux ordres d'enseignement, les conséquences reliées au partage des tâches ne sont que légèrement avantageuses (échelle d'évaluation) et à peine applicables alors que les autres conséquences sont de moyennement à fortement avantageuses et presque moyennement applicables.

En excluant la conséquence d'entraîner une augmentation de la tâche, les deux énoncés qui restent liés au partage des tâches (diminu-

tion du nombre d'élèves par groupe et du nombre de groupes par enseignants) obtiennent les cotes moyennes suivantes.

Ordre	Appartenance	Évaluation	Applicabilité
primaire	1,17	2,66	0,93
secondaire	1,47	2,83	0,82

Il s'agit d'énoncés de légèrement à moyennement conséquents, mais jugés très avantageux et moins qu'à peine applicables. Dans une étape suivante, six regroupements des conséquences ont été effectués selon leurs valeurs extrêmes (les plus fortement et les plus faiblement cotées) sur chacune des trois échelles. La construction de ces regroupements, le nombre de conséquences dans chaque groupe et leurs cotes sur l'échelle concernée, de même que les valeurs moyennes des cotes de ces conséquences sur chacune des échelles et pour chacun des ordres d'enseignement, sont décrits au tableau 6.

L'examen du tableau 6 appelle les constatations suivantes. En premier lieu, sur chacune des échelles d'appartenance et d'évaluation, les énoncés sont cotés semblablement par les enseignants des deux ordres d'enseignement, alors que sur l'échelle d'applicabilité, les enseignants du secondaire trouvent d'environ 0,25 moins applicables que ne le trouvent ceux du primaire les conséquences le plus fortement cotées sur chacune des échelles. En deuxième lieu, le maximum défini pour les échelles d'appartenance et d'évaluation est fixé à la valeur 3, alors que celui de l'échelle d'application est fixé à 4; pourtant, les cotes d'applicabilité maxima observées sont inférieures aux minima observées des cotes d'évaluation pour chacun des deux ordres et inférieures aux maxima observées des cotes d'appartenance pour les enseignants du secondaire.

Il convient de noter que, des huit conséquences les plus fortement cotées sur l'appartenance, deux font aussi partie des six conséquences les plus fortement évaluées, trois autres font aussi partie des sept conséquences les plus applicables, aucune ne fait partie à la fois des six plus fortement évaluées et des sept plus applicables et aucune n'appartient aux huit plus faiblement évaluées ni aux huit moins applicables.

Tableau 6 – Moyennes observées (M) sur les échelles relatives aux conséquences comportementales regroupées selon leur position extrême commune aux deux ordres d'enseignement sur ces échelles

Regroupement* des conséquences comportementales	(N)	Ordre d'ens.	Appartenance (0 à 3)	Évaluation (-3 à 3)	Applicabilité (0 à 4)
Les + fortement cotées sur					
Appartenance	(8)				
2,28 à 2,47		pr.	2,38	2,67	2,16
2,23 à 2,63		sec.	2,47	2,64	1,91
Évaluation	(6)				
2,72 à 2,87		pr.	1,90	2,78	1,97
2,62 à 2,87		sec.	1,92	2,78	1,70
Applicabilité	(7)				
2,15 à 2,65		pr.	2,10	2,61	2,36
1,92 à 2,233		sec.	2,16	2,63	2,11
Les + faiblement cotées sur					
Appartenance	(6)				
0,27 à 0,59		pr.	0,45	1,91	1,69
0,12 à 0,67		sec.	0,39	1,89	1,62
Évaluation	(8)				
-2,25 et 0,25 à 1,75		pr.	1,15	0,83	1,71
-2,48 et 0,88 à 1,93		sec.	1,16	0,90	1,43
Applicabilité	(8)				
0,84 à 1,61		pr.	1,48	1,57	1,28
0,77 à 1,34		sec.	1,54	1,55	1,14

* Pour chaque qualité (par exemple, les plus fortement cotées sur l'appartenance), les 10 conséquences qui y correspondent ont été identifiées pour chaque ordre d'enseignement séparément. N'ont alors été retenues pour le regroupement que les (N) conséquences qui appartiennent aux deux ensembles ainsi obtenus et dont l'étendue des cotes est décrite en colonne de gauche.

Outre les trois énoncés déjà mentionnés relatifs au partage des tâches, les huit conséquences jugées les moins applicables concernent le respect des rythmes individuels d'apprentissage, la fabrication et l'utilisation fréquente d'instruments appropriés de mesure ou d'observation, ainsi que la détermination de seuils de réussite pour l'interprétation critériée. Les aspects d'application difficile concernent donc les dimensions organisationnelles et d'instrumentation de l'évaluation formative.

6. CROYANCES SAILLANTES

Tel que cela a déjà été argumenté et annoncé, le regroupement des huit conséquences les plus fortement cotées sur l'échelle d'appartenance définit le domaine des croyances saillantes, décrit au tableau 7.

Ces huit croyances saillantes représentent les conséquences majeures, du point de vue des enseignants, de la pratique de l'évaluation formative. Toutes ces croyances concernent les interventions de l'enseignant : c'est à ce niveau que les enseignants situent l'impact majeur de la pratique de l'évaluation formative. Selon eux, cette pratique signifie que l'enseignant doit évaluer ses élèves de façon continue dans le but d'améliorer leur apprentissage (les 3 premiers du tableau 7), ce qui s'appuie sur l'identification des objectifs maîtrisés ou non (les 2 suivants du tableau 7) de même que sur le diagnostic et la correction des erreurs (les 3 derniers du tableau 7).

Cette perception correspond à la définition de l'évaluation formative énoncée par Scallon (1988, 175) et retenue pour l'étude. Elle supporte aussi la proposition déjà émise à l'effet que les enseignants manifestent une attitude favorable à la pratique de l'évaluation formative lorsqu'ils considèrent les bénéfices que retireraient les élèves des énergies et du temps qu'ils y investiraient.

Ces croyances saillantes, conséquences majeures de la pratique de l'évaluation formative, sont évaluées très positivement par les enseignants du primaire (2,56 à 2,75; moyenne de 2,67) et du secondaire 2,53 à 2,72; moyenne de 2,64). Il s'agit là de cotes dont la moyenne est à peine de 0,1 plus basse que la moyenne des 6 conséquences les plus fortement évaluées et dont font aussi partie les croyances saillantes 3 et 5 (tableau 7).

Tableau 7 – Les huit croyances comportementales saillantes*

Énoncés	Appartenance (0 à 3)		Évaluation (-3 à 3)		Applicabilité (0 à 4)	
À mon avis, pratiquer l'évaluation formative...	M**	S	M	S	M	S
amène l'enseignant à reprendre les explications qui n'ont pas été comprises par ses élèves (1)	2,47 2,63	1,08 0,70	2,72 2,72	0,67 0,65	2,65 2,83	0,89 1,05
amène l'enseignant à intervenir de façon continue pour améliorer l'apprentissage en cours (2)	2,43 2,63	0,94 0,75	2,69 2,64	0,70 0,66	2,15 1,92	0,86 101
oblige l'enseignant à observer de façon continue ses élèves en cours d'apprentissage (3)	2,47 2,49	0,81 0,99	2,72 2,62	0,56 0,69	1,90 1,77	0,93 0,95
permet à l'enseignant d'identifier chez chez ses élèves le degré d'atteinte de chacun des objectifs du programme (4)	2,39 2,55	1,231 0,95	2,61 2,56	0,70 0,92	1,95 1,77	0,85 0,96
Oblige l'enseignant à vérifier si les élèves possèdent les acquis nécessaires à la poursuite d'un apprentissage (5)	2,37 2,37	0,96 0,99	2,75 2,72	0,56 0,54	2,26 1,81	0,77 1,02
oblige l'enseignant à analyser les erreurs d'apprentissage commises par ses élèves (6)	2,37 2,37	0,93 1,00	2,66 2,64	0,75 0,79	2,08 1,86	0,79 1,11
oblige l'enseignant à modifier des interventions en fonction des types d'erreurs commises par les élèves (7)	2,30 2,46	1,15 0,95	2,56 2,68	0,80 0,73	2,34 2,20	0,84 0,99
oblige l'enseignant à faire suivre son diagnostic d'une action corrective appropriée aux difficultés d'apprentissage des élèves (8)	2,28 2,23	0,94 1,17	2,68 2,53	0,65 0,77	1,94 1,64	0,79 1,03

* Les valeurs en première ligne sont pour le primaire (Np = 129) et celles en seconde ligne sont pour le secondaire (Ns = 117).

** M = moyenne, S = écart type.

Cependant, l'examen des cotes d'applicabilité des conséquences qui définissent les croyances saillantes permet de constater que les enseignants jugent ces conséquences comme étant tout au plus moyennement applicables (de 1,90 à 2,65 avec une moyenne de 2,16 au primaire; de 1,64 à 2,33 avec une moyenne de 1,91 au secondaire), même si trois croyances saillantes (1, 2 et 7 du tableau 7) font aussi partie des 7 conséquences les plus applicables.

Trois des huit conséquences majeures qui servent à définir les croyances saillantes reçoivent une cote d'applicabilité inférieure à 2 et sont donc moins que moyennement applicables pour chacun des deux ordres d'enseignement; il s'agit des conséquences faire suivre son diagnostic d'une action corrective (nº 8), identifier le degré d'atteinte de chaque objectif (nº 4) et observer de façon continue (nº 3).

Ces conséquences qui sont pourtant hautement évaluées (jugées très avantageuses) concernent l'observation des élèves durant l'apprentissage.

CONCLUSION

Les croyances saillantes des enseignants des deux ordres d'enseignement s'appuient donc sur des conséquences majeures et pertinentes de la pratique systématique de l'évaluation formative, conséquences qu'ils jugent très avantageuses pour l'amélioration des apprentissages de leurs élèves, mais qui exigent aussi des conditions au plus moyennement applicables dans le contexte scolaire actuel.

Ces constats sont susceptibles d'aider à mieux comprendre ce qui motive (ou démotive) les enseignants à adopter diverses pratiques d'évaluation formative.

En premier lieu, les croyances saillantes identifiées semblent indiquer que les enseignants sont conscients de leurs rôles et de leurs responsabilités dans l'évaluation formative des apprentissages de leurs élèves. Le message conceptuel paraît donc assez bien transmis, du moins quant à la justesse des principales caractéristiques de l'évaluation formative et de sa fonction d'aider à améliorer les apprentissages des élèves.

Deuxièmement, les enseignants expriment un haut degré d'accord avec les avantages attribués aux divers aspects de l'évaluation formative.

Troisièmement, les enseignants mettent en question néanmoins l'applicabilité ou les conditions d'application des actions conséquentes attendues d'eux au nom de l'évaluation formative. Ce constat de faible applicabilité paraît donc être la meilleure explication d'un certain manque de ferveur des enseignants à s'engager dans des pratiques soutenues et systématiques d'évaluation formative des apprentissages de leurs élèves.

Les conditions d'application, qui restent à être recherchées et définies, concernent une dimension organisationnelle, rarement étudiées mais qui paraît incontournable. Une intention positive et une attitude favorable ne semblent pas suffire, même lorsqu'elles s'appuient sur des croyances comportementales saillantes justes et solides. C'est la dimension organisationnelle qui semble en cause pour une part majoritaire des objections soulevées par les enseignants qui voient leur marge de manœuvre en classe de plus en plus réduite.

Les principaux référents des enseignants en cette matière, appuyés d'une documentation de plus en plus abondante, ne peuvent dissimuler qu'opérationnaliser une pratique systématique de l'évaluation formative des apprentissages en classe, même en supposant que toute l'instrumentation nécessaire soit disponible à l'enseignant, risque de commander une gestion qui laisserait trop peu de place aux manœuvres de l'art d'enseigner. Si tel est le cas, la réticence des enseignants quant à l'applicabilité dresserait un mur de réalisme susceptible d'empêcher le motivation, l'attitude et l'intention de s'actualiser dans les pratiques.

Les référents (personnes, organismes, documents) en matière d'évaluation formative ont peut-être, à leur insu, trop embrassé ou trop visé, mais le balancier ne peut alors que frapper et rebondir sur le mur de la réalité de la classe. À titre d'exemples, malgré les énergies investies dans l'auto-évaluation des élèves, dans la confection de bulletins descriptifs, dans l'utilisation de tests diagnostiques, ce sont là, parmi d'autres, autant de conséquences jugées à la fois peu ou pas pertinentes, avantageuses et applicables par les enseignants. Vaut-il encore la peine de tenter d'intervenir pour les convaincre du contraire?

Ne vaudrait-il pas mieux reconnaître que les visées d'évaluation formative, transmises par les référents, ont atteint leur maximum réaliste en permettant que les enseignants d'aujourd'hui et de demain s'approprient les conséquences de l'évaluation formative sur l'amélioration des apprentissages de leurs élèves et en reconnaissent ainsi les grands avantages?

Ce maximum réaliste semble avoir déjà été atteint chez de nombreux enseignants, alors que beaucoup d'autres enseignants semblent encore imperméables à l'évaluation formative. En conséquence, plutôt que de placer l'accent sur la dimension organisationnelle de l'évaluation formative, d'autres énergies d'animation doivent être consenties pour assurer que les conséquences les plus pertinentes soient reconnues et jugées très avantageuses par tous les enseignants ou presque : un enseignant sur huit au primaire et un sur quatre au secondaire trouvent encore que pratiquer l'évaluation formative n'est pas bénéfique (voir tableau 3). Si l'on ajoute ceux pour qui cette pratique est tout au plus légèrement bénéfique, c'est de 30 % à 40 % des enseignants qu'il faut encore animer ou convaincre.

Comme constante de l'étude rapportée, il convient de souligner que, sur les variables intention, attitude, attentes perçues des référents, motivation à s'y conformer, applicabilité des conséquences comportementales, donc sur toutes les variables sauf les échelles d'appartenance et d'évaluation des conséquences, les enseignants du secondaire expriment plus de réticence (cotent plus bas) que ceux du primaire. Cette plus grande imperméabilité du secondaire peut-elle être attribuable à une difficulté de pratiquer l'évaluation formative réellement plus grande au secondaire qu'au primaire? Ou est-ce le résultat de moins d'animation dans les écoles secondaires que dans les écoles primaires?

Avant de poursuivre l'animation, il faut cibler les endroits, les objectifs, les motifs de résistance. L'étude rapportée ici vise à contribuer à éclairer le débat, les besoins, les décisions, les actions. Ces dernières, pensons-nous, doivent être orientées vers le développement et la promotion de l'évaluation formative davantage en termes d'attitudes et de conduites d'enseignement qu'en termes de conduites d'évaluation.

En effet, le temps et les énergies consacrés à l'évaluation risquent d'empiéter sur le temps et les énergies nécessaires pour réaliser les tâches d'enseignement.

Quant à l'instrumentation qui doit supporter l'évaluation formative, non seulement elle doit être adéquate et fournie ou rendue disponible aux enseignants, encore faut-il qu'elle soit conçue de manière à supporter une rétroaction pertinente et la plus immédiate possible sans imposer aux enseignants un surcroît de tâches, et de manière à ne pas accroître indûment le temps de classe consacré à l'évaluation, ce qui restreindrait celui consacré aux apprentissages.

RÉFÉRENCES

AJZEN, I. et FISHBEIN, M. (1980).
Understanding attitudes and predicting social behavior. Toronto : Prentice-Hall.

BURELLE, R., GADBOIS, L., PARENT, C. et SÉGUIN, S.P. (1991).
Mesurer des croyances et des attitudes pour mieux comprendre des comportements. *Mesure et évaluation en éducation, 14*(1), 61-77.

FISHBEIN, M. et AJZEN, I. (1975).
Belief, attitude, intention and behavior : An introduction to theory and research. Reading (MA) : Addison-Wesley.

GADBOIS, L., BURELLE, R., PARENT, C. et SÉGUIN, S.P. (1991).
Un instrument de mesure des croyances et attitudes des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves. *Mesure et évaluation en éducation, 14*(2), 5-24.

PARENT, C., SÉGUIN, S.P, BURELLE, R. et GADBOIS, L. (1991).
Croyances et attitudes des enseignants de l'ordre primaire de la région administrative du grand Montréal métropolitain à l'égard de la pratique de l'évaluation formative des apprentissages de leurs élèves. *Actes du colloque de l'AFIRSE, (Association francophone internationale de recherche scientifique)*, Carcassone.

PARENT, C., SÉGUIN, S.P. et BURELLE, R. (en préparation).
Utilité du modèle de Ajzen et Fishbein pour mesurer les croyances et les attitudes en éducation.

SCALLON, G. (1988).
L'évaluation formative des apprentissages (Tome 1 - *La réflexion*). Québec : Les Presses de l'Université Laval.

L'ACTE D'ÉVALUER LES APPRENTISSAGES : PROBLÉMATIQUE DE LA FORMATION

LOUISE BÉLAIR
Université d'Ottawa

INTRODUCTION

L'acte d'évaluer au primaire comme au secondaire a fait l'objet de multiples débats et de multiples congrès au cours des vingt dernières années, surtout en ce qui concerne l'évaluation formative et son application en salle de classe. Cependant, il demeure un malaise chez l'enseignant ou l'enseignante, car cet acte mystérieux est encore empreint de valeurs et de peurs. Il en est souvent question, quoique sans trop de profondeur, car il reste trop d'imprévus et de facteurs à gérer lorsque vient le temps d'agir. Les multiples formations données par le biais de journées pédagogiques ou par les cours de formation continue offerts par les universités tentent d'amener les participants et les participantes à mieux définir et à mieux gérer leur action d'évaluation. C'est dans une perspective plutôt provocatrice que s'insère ce texte afin d'amener le lecteur ou la lectrice à se positionner face à sa propre démarche d'évaluation ou à la démarche de formation continue qu'il ou elle entreprend.

La première partie de ce propos définit l'acte d'évaluer, vu du point de vue de la personne évaluatrice, comme un geste qui implique diverses habiletés. Dans la perspective du milieu scolaire, c'est d'abord de savoir où nous en sommes, où nous désirons aller et surtout comment y arriver qui permet de situer plus clairement l'action d'évaluer. Ainsi, la conception de l'action d'évaluer se présente comme un acte humain intériorisé où le geste posé, qu'il soit réfléchi ou non, qu'il soit positif ou négatif, implique une prise de position envers un autre individu ou envers une situation et oblige le développement d'habiletés qu'elles soient implicites ou explicites. L'analyse de ce geste est aussi présentée en parallèle avec le type de formation que reçoivent actuellement les enseignants et les enseignantes.

Dans un deuxième temps, à la suite de cette analyse et en fonction de certaines recherches effectuées sur la problématique du changement

et sur l'apprentissage, il s'agit de cerner les étapes de formation qui permettent un véritable changement de comportement et d'attitude en matière d'évaluation. On pourrait alors penser viser une déconstruction du concept qui impliquerait une déformation du vécu actuel des intervenants et des intervenantes dans le milieu. La formation qui en découlerait permettrait, entre autres, une sensibilisation aux innovations, aux expériences extérieures, aux organisations et aux systèmes de gestion possibles pour évaluer judicieusement les apprentissages. De tout ceci devrait découler une transformation de l'individu dans ses actions pédagogiques et évaluatrices, à la suite de cette transaction entre la déformation et le savoir expérientiel.

1. L'ACTE D'ÉVALUER AU PRIMAIRE : UNE ACTION HUMAINE INTÉRIORISÉE

Quoiqu'il soit plutôt difficile de dissocier les rôles de l'enseignant ou l'enseignante dans ses multiples tâches reliées à la salle de classe et aux élèves, il demeure qu'il s'agit de cerner le rôle particulier qui entraîne l'évaluation des apprentissages, quelle soit formative ou sommative. Cet acte exige des habiletés plus complexes que l'acte d'enseigner, de gérer une classe ou d'intervenir en sous-groupes, car il demande de juger des actions et des faits extérieurs à sa personne et stipule une relation entre plusieurs individus.

1.1 L'acte d'évaluer implique un jugement moral complexe

L'étude de diverses théories, tant en psychologie développementale qu'en sociologie, porte à croire que l'évaluation est une habileté fort complexe qui exige des valeurs hautement morales et une fonction élevée de jugement. Reboul (1992) donne cinq traits relatifs au jugement, à savoir qu'il implique un monde de relations humaines, qu'il est une décision créant une solution, qu'il engage et compromet l'acteur, qu'il doit motiver la décision et qu'il est moral. Si l'on identifie l'évaluation comme une appréciation envers autrui, on la définit comme un jugement et l'on doit alors déterminer l'action d'évaluer en fonction de ces cinq traits. Même si l'acte peut être irréfléchi ou plutôt spontané à la suite d'une situation qui exige une prise de position immédiate, il demeure qu'il exige de l'actant une décision consécutive à une réflexion (quoique courte) et surtout un engagement.

De son coté, Bloom (Bloom, Hasting et Madaus, 1971) détermine l'évaluation comme l'habileté la plus complexe du domaine cognitif en stipulant qu'évaluer implique un jugement porté à partir de critères élaborés au préalable, pour une critique interne et/ou externe de l'objet évalué. Cette évaluation, perçue de manière plus idéalisée, devrait alors se faire par un expert dans un domaine, en fonction des règles régissant ce domaine, par rapport à un but déterminé. Ainsi, le concept d'évaluation, par sa fonction et sa définition (Cardinet, 1986; Scallon, 1988), imposerait certains principes à l'évaluateur, tels accorder un jugement objectif et assurer une équité envers tous les évalués.

Kohlberg (1969) indique que l'équité représente le stade de jugement moral le plus élevé, le sixième stade, dans la hiérarchie des pensées et qu'il se situe au niveau postconventionnel dans la structure de jugement moral. Par ailleurs, certaines recherches recensées par l'auteure (Bélair, 1990) ont démontré que peu d'individus atteignent ce niveau de pensée et qu'il est probable que la majorité des types d'enseignement se retrouvent surtout au quatrième stade du jugement moral : respect de l'ordre et de l'autorité. Ceci pourrait s'illustrer dans l'analyse des pédagogies traditionnelles où l'on constate que l'enseignement se conjugue par une autorité suprême de la classe, ayant pouvoir sur les élèves, portant un regard critique et contrôlant sur les productions, et assurant une discipline unique (Perrenoud, 1988). Les pédagogies actives ont tendance à définir la relation enseignant/élève selon des schèmes de pensée qui réfèrent plutôt au cinquième stade de Kohlberg (le bien-être et la démocratie) où l'élève est traité à égalité et où le respect de l'autre est primordial.

L'évaluation prise à part, on retrouve alors un acte de jugement envers autrui, donc déterminé par le sixième stade de Kohlberg, l'équité. Ainsi, peu importe l'approche pédagogique favorisée en salle de classe, il demeure que l'acte d'évaluer exigerait une structure de pensée différente de l'acte d'intervention auprès des élèves, quel que soit ce type d'intervention. L'enseignement exige alors une constante oscillation entre l'intervention pédagogique et l'action d'évaluer, entre sa propre conception de l'enseignement et celle de l'évaluation, soit entre deux stades de jugement moral, sautant parfois du quatrième au sixième stade.

Ainsi, il semble logique que cet acte se matérialise en des actions hautement subjectives et surtout hautement périlleuses pour l'évalué(e). Même si l'évaluateur ou l'évaluatrice possède la technique de construction d'examens la plus sophistiquée, même s'ils effectuent les analyses statistiques les plus poussées, il demeure que, dans son intention et dans sa forme, l'évaluation reste une action humaine qui atteint les hauts niveaux taxinomiques des échelles socio-affectives (Krathwool, 1976). La personne évaluatrice consciente de son rôle doit bien se connaître en tant qu'actant dans ce processus. Il faudrait être conscient de l'importance de l'équité lorsque vient le moment de juger de ses acquis; une équité centrée sur l'évalué(e) et non sur le groupe, sur le programme, sur la direction ou sur les parents demanderait une attention constante envers chacun des individus qui dépendent de cette évaluation. C'est un contexte de pédagogie différenciée, ou plutôt d'évaluation différenciée.

De l'action humaine et spontanée à l'action réfléchie et organisée, il se peut que la marge soit grande et plutôt insécurisante pour l'actant. Ce dernier est souvent conscient de ses lacunes comme évaluateur et se sent démuni face à des décisions rapides qu'il doit prendre. Les principes d'équité, de justice, d'action valorisante et de respect envers les autres ne sont pas toujours facilement applicables.

1.2 L'acte d'évaluer est différent selon l'évaluateur

La personne qui évalue une situation, d'autres personnes ou un fait devrait cependant faire le point sur ses propres valeurs et sur ses propres craintes face à cet acte. L'évaluation exige une intériorité de sa personne. En effet, il ne s'agit pas simplement de mesurer l'impact des décisions sur l'avenir de l'élève, ou d'essayer d'intégrer la démarche d'évaluation à la situation d'apprentissage, il faut s'approprier l'acte et y arrimer sa propre définition (Robillard, 1990). L'action d'évaluer ne saurait être mécanique ou technique, mais plutôt une résultante de ses valeurs dans une démarche humaine.

«La personne qui évalue le fait avec l'ensemble de ses habiletés intellectuelles, de ses impressions, de ses perceptions, de son expertise, de son intuition, de sa subjectivité, de ses sensations, de ses sentiments, de ses expériences,... de ses valeurs, de ses tensions, de ses conflits, de ses ambiguïtés, de ses incertitudes, de ses malaises» (Robillard, 1990).

La personne enseignante, dans son rôle d'évaluatrice, devrait d'abord pouvoir analyser son vécu intérieur et développer une harmonie entre ses convictions profondes et ses actions, indépendamment des demandes extérieures, des programmes, de l'administration et des parents. Selon Angers (1992), une telle habileté complexe exige une appropriation de soi afin de cerner le niveau de conscience intentionnelle qu'on y accorde. Qu'il s'agisse d'expérimenter, de comprendre, de juger ou de décider, l'enseignant ou l'enseignante devrait développer un processus de réflexion sur ses actions de jugement et de décision avant même de les effectuer.

Cette introspection renforce le fait que l'acte d'évaluer est complexe, que l'habileté est élevée, que le jugement moral est centré sur l'équité et que l'actant doit développer une action humaine plutôt que mécaniste de l'évaluation. Mais la complexité de la tâche augmente par les pressions de la société qui oblige certaines orientations semblant valoriser soit des actions non conformes à son intériorité, soit des actions dont on ne suppose pas l'effet néfaste chez lui ou chez l'élève.

Les enseignants sont de plus en plus sollicités par des technologies nouvelles, par des méthodologies qui mettent en «péril» leur pouvoir de maître, par des programmes et des stratégies de plus en plus prescriptifs qui, par le fait même, leur demandent de faire de plus en plus de choix. Beaucoup ne se sentent pas outillés devant ces choix. Le manque de temps, de ressources, de matériel, de connaissances et d'appui sont souvent les causes majeures externes qui peuvent justifier un certain manque de confiance dans les choix à effectuer, tant en évaluation qu'en enseignement. Le plus souvent, on retrouve des personnes qui utilisent un vocabulaire approprié, des concepts innovateurs, des habiletés centrées sur la personne et l'équité, mais qui ne mettent pas en pratique ces idées, ces concepts et ce vocabulaire désirablement acceptés et véhiculés par la société (Dassa, 1992). De plus, l'écart entre le moment où ils ou elles ont reçu la formation de base et les problématiques actuelles devant être mises en place est si grand, qu'il semble infranchissable à plusieurs intervenants et intervenantes en milieu scolaire.

L'évaluation rappelle des souvenirs quelque peu négatifs chez la plupart des personnes. Beaucoup ont connu des problèmes d'injustice

ou de manque de rigueur dans une correction. D'autres identifient leurs craintes devant des situations hautement stressantes qu'elles classent comme des actes répressifs d'évaluation. Ainsi, ce sentiment revient à la surface lors de l'évaluation lorsque l'actant a tout-à-coup peur de ne pas bien faire, de se tromper et de générer une situation peu équitable. Cette peur s'accentue par un sentiment d'incompétence devant toutes les nouveautés et provoque alors un recul devant le changement que l'on voudrait amorcer. D'ailleurs, les diverses formations continues peuvent accentuer ce recul, plutôt que le diminuer, en rappelant constamment cette impression négative de soi et d'incompétence.

1.3 L'acte d'évaluer est différent selon l'objet

L'objet de l'évaluation n'est pas unique, ce qui a tendance à complexifier encore plus cette action. Cardinet (1986) identifie trois subdivisions possibles des types d'apprentissage : les contenus, les démarches de pensée et les fonctions du savoir en situation. Leur évaluation nécessite des habiletés différentes de la part de l'évaluateur. On exige une expertise au niveau du contenu, une conscience des divers mécanismes de pensée afin de mieux intervenir auprès des élèves et une vigilance constante, car l'apprentissage et l'évaluation se font «en action» et que l'on doit établir toutes les formes de progression possibles.

Cazabon (1990) détermine à son tour trois domaines d'apprentissage : savoir penser, savoir communiquer et savoir s'actualiser. Quoique les deux premiers domaines s'apparentent avec les catégories de Cardinet, il demeure que l'actualisation ajoute à la complexité de l'acte d'évaluation. Évaluer l'actualisation de soi chez l'élève oblige l'évaluateur à circonscrire ce domaine, ce qui impliquerait alors une appropriation de ses propres valeurs, de son propre cheminement et de ses propres croyances (donc intérioriser l'acte selon Robillard), en vue de faire des choix judicieux pour lui-même. Il faudrait donc avoir déjà fait cette démarche avant de l'identifier chez les autres et de les aider par le biais de l'évaluation. De Ketele (1992) renforce ce phénomène en identifiant les savoirs devenir et les savoirs d'expertise de l'apprenant. Cette dimension ajoute à la complexité, car il s'agit de tenir compte du lien entre les connaissances antérieures des élèves et les nouveaux contenus abordés avec eux (Tardif, 1992).

Ces découpages théoriques des objets d'évaluation sont plus ou moins définis dans les programmes d'études ou dans les programmes cadres des divers organismes scolaires. On s'attend que l'enseignant ou l'enseignante découpe son programme et introduise une logistique personnelle de l'organisation des apprentissages à réaliser en une année scolaire. Ceci exige le développement d'une expertise dans tous les domaines enseignés, afin d'assurer une appropriation d'un curriculum qui corresponde à ses propres choix et à ceux de ses élèves.

1.4 L'acte d'évaluer est différent selon l'évalué

Pour l'élève, la réalité semble peu engageante. L'évaluation est, depuis toujours, source de peur, de conflits, de négativisme. C'est une source intensive de stress, un regard de l'autre sur son vécu, une appréciation plus ou moins recherchée sur ses performances. L'évaluation permet de sanctionner, de graduer, de féliciter, de détruire, de punir. Toutes ces actions incluses dans l'acte d'enseigner font en sorte que la réaction de l'évalué(e) est plutôt défensive et négative. L'évalué(e) se protège en situant le contexte d'évaluation comme «un mal nécessaire» ou une «façon qu'a le professeur de nous dire qu'on est nul», toute forme de protection étant centrée sur des causes extérieures à son propre cheminement. L'évaluation est aussi source de promesses, de cadeaux, de plaisir... à la condition de réussir, de «faire mieux», d'être à la hauteur des espérances des autres, c'est-à-dire des parents, des adultes. Que le formulaire de bulletin du primaire soit maintenant descriptif, qualitatif, avec commentaires, sans comparaison avec le groupe classe ne semble rien changer à la situation, ou si peu! Il est évident que les résultats sont moins quantifiés et moins comparés, mais les valeurs inhérentes à l'évaluation ne sont pas nécessairement changées pour autant chez l'évalué(e), que chez l'évaluateur ou l'évaluatrice; cela fait partie du jeu du contrat didactique.

Par ailleurs, les chercheurs parlent de plus en plus de pédagogie différenciée, centrée sur les élèves, en fonction de leurs besoins (Scallon, 1988; Paré, 1977). Ils mentionnent aussi que l'acte d'évaluer va dans ce sens, s'il est formatif et s'il répond à l'aide demandé par l'élève en apprentissage (Perrenoud, 1992). La personne dans ce rôle d'évaluateur devrait pouvoir déterminer les causes et les effets des échecs de

chacun des élèves, y remédier par des activités qui répondent à ces besoins et en assurer un cheminement le plus personnel possible, pas nécessairement uniforme pour la classe, vers la maîtrise des apprentissages de chacun (Allal, 1992). Cette spécification de l'acte d'évaluer entraîne une refonte complète dans la pédagogie traditionnelle et détermine un contrat didactique différent du contrat traditionnel, où la relation élève/maître s'avère changée, devient démocratique et rejoint alors la description du cinquième stade de Kohlberg. Ainsi, un tel effort de réflexion sur l'acte d'évaluer entraînerait chez l'enseignant ou l'enseignante une redéfinition de son rôle qui va beaucoup plus loin que l'action d'évaluer et surtout exigerait un changement majeur de ce rôle. Dans une telle perspective, on ne peut pas décrire l'acte d'évaluer comme un acte unique, identique et systématique pour tout un groupe d'élèves, ni comme un acte sans conséquence pour l'enseignement, mais plutôt comme un acte différencié selon les contraintes, la clientèle, les diverses difficultés des élèves, les moments privilégiés, tout en étant équitables et tout en respectant une technique précise d'élaboration.

C'est ainsi que l'on retrouve divers paradoxes entre le vécu des enseignants et des élèves et entre ce qu'on exige d'eux. Perrenoud (1991) dénote trois grands paradoxes attribuables à la complexité des tâches d'enseignement et d'apprentissage, dont le troisième touche particulièrement les nouvelles tendances en évaluation pédagogique : la pédagogie de la maîtrise aussi définie comme la pédagogie de la réussite (Huberman, 1988). Pour aider l'élève en difficulté et lui rendre le succès accessible, il faut lui signifier ses faiblesses et lui donner la chance de les surmonter. C'est l'évaluation formative. Il est alors possible que l'énumération des faiblesses d'un élève, même dite positivement, puisse engendrer un découragement et une désillusion de la part de cet élève devant la montagne de difficultés auxquelles il a à faire face pour réussir. Ainsi, l'acte d'évaluer, même centré sur l'élève, devrait tenir compte du fait que l'élève, comme l'enseignant ou l'enseignante, peut naviguer entre l'état de découragement et l'illusion que tout va bien, et tenir compte que ce dernier peut simplement avoir le goût de «faire l'autruche».

1.5 L'acte d'évaluer est avant tout interactif

La recherche actuelle parle de l'évaluation comme d'un acte de communication et de relations entre deux ou entre plusieurs personnes. Ouellette (1990) voit l'acte d'évaluer comme un processus dynamique et non comme un élément isolé. L'évaluation est «...la synthèse de l'expérience observable, en faisant accéder les individus aux diverses dimensions que renferme tout phénomène éducatif» (Ouellette, 1990). Toujours selon Ouellette, l'évaluation permet le dialogue entre les individus et la situation, elle permet d'établir les caractéristiques d'une situation éducative et enfin assume un ensemble de règles ou de lois auxquelles se soumettent les «communicateurs». Perrenoud (1991) soutient cependant que toute communication n'est pas nécessairement positive, logique, essentielle et synonyme de coopération. En effet, elle peut être répressive, trompeuse, manipulatrice, hiérarchique, aliénante et inconsciente.

Que l'évaluation soit sommative ou formative, elle peut donner une chance à l'élève; une chance de réussir ou une chance de tricher. Ces actes d'évaluation pourraient être alors pris comme des communications tantôt confiantes, tantôt séductrices. Ces communications sont d'autant plus difficiles qu'elles se vivent en situation de classe, avec 25 ou 30 autres personnes en constante cohabitation. Sirota (1988) explique qu'on peut passer d'un réseau officiel à un réseau plus clandestin dans les divers moments d'évaluation. Allal (1988, 1992) parle d'interactions maître-élève, entre élèves et élève-matériel qui se définissent comme des réseaux officiels de communication d'évaluation. La clandestinité fera référence aux attitudes, au non-verbal, aux commentaires impromptus et aux non-réponses aux questions. Ces actes peuvent être joués correctement ou non selon le milieu socioéconomique, les types de stratégies utilisées et les acteurs en relation (Sirota, 1988). Par ailleurs, toute pratique a un langage qui va de l'action réelle à l'action idéale vue de l'actant (Robillard, 1992); ceci implique alors que l'enseignant ou l'enseignante devrait s'approprier son propre langage verbal ou visuel, avant de poser un geste déterminant pour l'élève.

1.6 Une réalité ... troublante!

Même si les tendances politiques et ministérielles des années quatre-vingt et début quatre-vingt-dix ont donné à penser que l'évaluation fait partie de l'apprentissage et qu'elle doit avant tout permettre à

l'élève d'atteindre un niveau de réussite plutôt que d'aspirer à devenir le meilleur du groupe; même si le passage d'une évaluation à interprétation normative à une interprétation critérielle basée sur la pédagogie de la maîtrise semble évident pour plusieurs (MEQ, 1984, 1985), il s'avère que l'action entreprise est assez loin des principes dégagés sur papier.

La réalité de la classe représente une personne responsable d'un groupe d'élèves tous différents. Cette personne a pour mandat de fournir toutes les occasions possibles d'apprentissage à chacun de ces élèves afin que tous réussissent selon des exigences minimales. Cette personne doit alors informer les parents et les administrateurs des seuils de réussite de l'élève en fonction des divers apprentissages. Cette personne, selon ces divers rôles, est donc à la fois le «maître» qui juge et entérine, «l'ami» qui confronte et aide, et «le juré» qui donne son verdict au pouvoir extérieur.

Après avoir dressé une contextualisation de l'action ou du geste posé par l'évaluateur, il semblerait que cet acte d'évaluer est beaucoup plus complexe que le laisse croire les quatre étapes de la démarche d'évaluation telle qu'elle est définie par le ministère de l'Éducation du Québec (MEQ, 1984). Au-delà de ces quatre étapes, cet acte, qu'il soit réfléchi ou spontané, qu'il soit positif ou négatif, demanderait idéalement de la personne évaluatrice : une connaissance approfondie d'elle-même, une capacité d'entrer en relations avec les autres, une expertise dans le champ d'étude, une étude psychologique du fonctionnement de la pensée, une capacité d'observer l'élève dans l'action entreprise, un jugement équitable et respectueux de l'autre et une appropriation des techniques en évaluation.

La pratique de l'évaluation en classe demeure très éloignée des croyances et des représentations que s'en font les enseignants et les enseignantes. Ces derniers mettent les concepts sur des actions qui ne sont pas toujours reliées à ces concepts. Même s'il y a eu un immense pas de fait dans les pratiques d'évaluation des enseignants ou enseignantes et même si les recherches actuelles permettent l'émergence de pratiques et de gestion nouvelles, il demeure que les valeurs inhérentes à tout acte d'évaluation semblent escamotées dans la quotidienneté des urgences, des imprévus, des «feux à éteindre» à longueur de journée.

La tâche d'enseignement relève de celle d'un gestionnaire du risque ou plutôt d'un «pompier».

Tout cet écart entre la réalité et les politiques ministérielles remettent en question l'apport des méthodologies nouvelles en évaluation des apprentissages. Plusieurs enquêtes journalistiques et plusieurs opinionnaires ont fait ressortir, à l'automne 1991, l'urgence de rétablir une situation qui semble catastrophique. Certaines réalités telles un bulletin descriptif utilisé comme un bulletin chiffré transformé en légende; des devoirs, des évaluations formatives et des exercices comptés pour insérer un résultat au bulletin; des opinions telles «l'évaluation sert à tranquilliser les parents, à satisfaire la direction», ou tout simplement un rappel du nombre incessant de directives plus ou moins contradictoires, de politiques plus ou moins logiques, de surplus de travail et d'une augmentation de la tâche (heures/contact élèves ou nombre d'élèves par classe) sont autant de facteurs qui restreignent l'innovation et la perspective d'un changement. On constate alors que cette situation n'encouragera pas non plus un changement des mentalités chez l'élève qui pense encore que l'évaluation est stressante, qu'elle est injuste et qu'elle est négative.

1.7 La formation actuelle au Québec et en Ontario

La formation à l'évaluation a cependant évoluée depuis les deux dernières décennies. Le passage du découpage des objectifs à une évaluation plus centrée sur les apprentissages complexes et le passage de l'évaluation bilan à une évaluation en situation sont autant de preuves d'une amorce d'évolution dans le champ de la recherche en évaluation.

Par ailleurs, on remarque un malaise causé par le manque de formation adéquate en évaluation lors de la formation initiale en enseignement car, même si ce cours est obligatoire (ce qui n'est pas le cas au Québec)[1], il est trop court (30 à 45 heures) et il est trop centré sur l'aspect technique de construction d'objectifs et d'instruments d'évaluation. En outre, compte tenu que ce cours est pratiquement le seul que les enseignants et les enseignantes vont suivre dans leur carrière (la formation continue universitaire n'étant pas obligatoire), les professeurs et profes-

[1] Synthèse des propos recueillis lors du symposium sur les «exigences minimales en évaluation dans la formation initiale des maîtres» tenu le 11 mai 1992 à Montréal dans le cadre de l'ACFAS.

seures ont tendance à vouloir donner le maximum d'informations, sans approfondir les prémices.

Pour sa part, la formation continue se situe à deux niveaux : les programmes des premier, deuxième et troisième cycles de formation continue offerts par les universités et les journées de perfectionnement organisées par le système scolaire. Ce perfectionnement peut être orienté en fonction des besoins manifestés par les enseignants et les enseignantes d'une école donnée. Cependant les budgets restreints et les orientations privilégiées de l'administration font en sorte que ces derniers sont souvent donnés à tous les enseignants et enseignantes d'une région, avec plus ou moins de préoccupations de leurs besoins manifestes. Dans ce cadre, certaines journées sont consacrées à la démarche d'évaluation, mais il demeure que ces rencontres semblent plutôt superficielles et amènent peu de changement réel.

Ces perfectionnements mettent en évidence certaines contradictions du système actuel. En effet, on peut vouloir donner tous les perfectionnements possibles et y croire, mais ces derniers peuvent rester inefficaces. Ainsi, le problème pourrait s'illustrer comme une manière de présenter le concept de l'acte d'évaluer pour amener un individu à changer sa vision et par le fait même sa démarche. Les ateliers de journées pédagogiques semblent à toutes fins presque inutiles, car la gestion des changements devient imposée et entre évidemment en contradiction avec le fait que l'on vise un changement (Perrenoud, 1989). Il faudrait envisager la formation continue sur une base de gestion du changement planifié (Savoie-Zajc, 1990).

2. ÉVALUER LES APPRENTISSAGES DES ÉLÈVES : UNE QUESTION D'APPRENTISSAGE

Il est difficile de cerner un processus d'apprentissage et de formation à un acte aussi complexe que l'évaluation. Il est évident que les personnes impliquées dans un processus d'évaluation n'ont pas les mêmes forces, ni les mêmes expertises, ni les mêmes façons d'apprendre. Il faut donc essayer de percevoir le niveau d'action de chaque intervenant et de prévoir avec eux un protocole de formation qui vise à long terme un changement dans leurs comportements, dans leurs attitudes et dans leurs valeurs.

2.1 L'évaluation perçue encore comme une innovation

L'enseignant ou l'enseignante a développé un ensemble d'idées préconçues face à l'évaluation. Ces préjugés sont pour la plupart des méconnaissances des théories sous-jacentes aux techniques d'évaluation proposées, des barrières devant une montagne de nouveautés difficiles à assimiler en peu de temps et des croyances populaires qui remontent à quelques dizaines d'années lorsque ces derniers étaient évalués.

Les croyances tels qu'elles sont énumérées au tableau 1 se dégagent de diverses analyses sur les croyances et les représentations des enseignants ou enseignantes face à l'évaluation (Fontaine et Trahan, 1990; Van Nieuwenhovan, 1990).

Tableau 1 — Inventaire partiel de croyances face à l'évaluation

Quant à l'élève	L'évaluation (les notes) sert à motiver les élèves à travailler. L'élève veut toujours savoir s'il est le meilleur. Seuls les aspects cognitifs peuvent être évalués; on ne peut pas évaluer les attitudes. Les élèves ne travaillent que si cela compte.
Quant à la gestion de classe	L'évaluation formative oblige à passer trop de temps à évaluer et moins de temps à enseigner. L'évaluation doit s'effectuer de la même manière pour tous et à tous au même moment.
Quant au système	Selon les administrateurs, la moyenne de la classe doit être entre 65 et 75 %, donc il ne sert à rien de différencier l'enseignement et de faire de l'évaluation formative. Les examens certificatifs des commissions scolaires visent surtout la comparaison entre les écoles sur la performance des élèves.
Quant aux parents	Les parents veulent des notes et savoir si leurs enfants sont les meilleurs. Les parents veulent un grand nombre de notes cumulées en cours de session.
Quant aux outils d'évaluation	Les examens à questions fermées sont plus objectifs et les questions ouvertes engendrent une subjectivité à dénoncer. Il est impossible d'observer en classe. Les enfants sont trop jeunes pour s'autoévaluer.

Ces croyances sont le bilan de plusieurs réflexions faites par des enseignants et des enseignantes québécoises et franco-ontariennes, lors de formations continues en évaluation pédagogique des apprentissages, données au cours des quinze dernières années.

Cet ensemble d'opinions reflète l'état conceptuel dans lequel se retrouvent les personnes directement impliquées dans ce processus. L'écart entre ces croyances et les pédagogies plus actives pratiquées en classe démontre alors un malaise dû au fait qu'il y a peu de cohérence entre les deux rôles d'enseigner et d'évaluer. L'évaluation semble perçue comme un concept isolé de ceux de l'enseignement et de l'apprentissage et se présente comme une innovation qui n'a pas encore trouvé sa véritable place dans le système de la classe. En effet, l'enseignant ou l'enseignante, comme l'élève, ne voit pas sa classe comme une organisation sociale où l'action d'évaluer s'inscrit dans les deux sens (élève et enseignant) en mettant les acteurs en relation symétrique en fonction d'un objet donné. Ces acteurs en situation d'évaluer semblent plutôt voir la classe comme un environnement de pouvoir dans lequel le patron décide et l'employé agit et, par le fait même, se déconnecte de toute situation d'apprentissage. Dans un tel contexte, il semble que l'évaluation «interactive» devienne une grande source de stress et d'incohérence, car elle ne correspond pas à l'organisation sociale en situation d'évaluation, même si elle y correspond lors des apprentissages.

2.2 Apprendre à évaluer, c'est comme apprendre à lire!

Les théories de l'apprentissage (Rogers, Bruner, Angers, Gagné, etc.) déterminent les diverses composantes de l'acte d'apprendre pour l'élève. Qu'on soit davantage cognitiviste, constructiviste, structuraliste ou non directiviste, que le climat de la classe soit davantage autoritaire, permissif ou coopératif, que les stratégies développées visent plutôt le cerveau droit ou le gauche, qu'elles soient plus visuelles ou que les interventions soient plus de nature coopérative, il demeure que l'enseignant ou l'enseignante est aux prises avec une foule de directives, d'orientations, de choix et d'organisations différentes qui l'influencent dans sa prise de position en classe.

Ces diverses théories sont détaillées, précisées et peaufinées selon les didactiques, selon les problématiques dues à certains contextes, selon la clientèle et selon le milieu socioéconomique. L'apprentissage de la lecture a particulièrement fait l'objet de modes et de théories en passant du «sablier» à la «méthode globale», sans en oublier toutes les ramifications. Ainsi, selon Artaud (1989), l'adulte est aux prises avec celles-ci (même parfois devenues des modes) et doit établir son propre cheminement en tentant d'établir une manière de «transmettre autrement» par le biais de la reconstruction et de la redéfinition des valeurs. Alors, même si la recherche a traité du lien entre ces théories d'apprentissage et l'évaluation des élèves (MEQ, 1984; Scallon, 1988; Allal, 1988), il demeure que ce lien est peu explicite dans les formations à l'évaluation.

On peut cependant prétendre qu'apprendre à évaluer tient du même phénomène que l'apprentissage des élèves. Doyon et Juneau (1991) expliquent comment faire apprendre l'autoévaluation aux élèves, en donnant les étapes d'un projet de formation que l'enseignant ou l'enseignante peut entreprendre en classe. Il serait alors envisageable d'apprendre à l'adulte comment faire cette autoévaluation sur lui-même et sur ses actions, avant d'entreprendre une évaluation des apprentissages des élèves. Les formations actuelles semblent oublier les prémices et les orientations des théories de l'apprentissage. Les vitesses d'exécution et les demandes rapides de résultats tangibles font en sorte que la formation actuelle ne répond pas aux postulats des théories de l'apprentissage.

Artaud amorce l'apprentissage comme une redéfinition des concepts. Dans la formation continuée, on pourrait parler de déconstruction de concepts, car les enseignants ou enseignantes ont un ensemble d'opinions dont il faut tenir compte à propos de l'intégration de l'évaluation dans l'apprentissage et un savoir expérientiel qui alimente leurs actions et leurs réflexions. Cette déconstruction doit se réaliser avec les participants et les participantes, car elle ne sera efficace qu'au moment où il y aura prise de conscience de ses propres actions par l'individu. Il s'agit de construire ensemble (formateur et enseignants) un nouveau concept et une nouvelle orientation. Ceci amènera alors automatiquement une transformation dans l'action et une appropriation interne de cet acte humain.

Si le phénomène de l'apprentissage est le même chez l'adulte que chez l'enfant, il faut alors se référer aux théories de l'apprentissage, définir les contenus en cause, le motiver et lui permettre de se réaliser dans l'action. Apprendre à évaluer doit donc toucher aux points développés antérieurement (sous partie 1.6), à savoir une connaissance approfondie de soi, une capacité d'entrer en relations avec les autres, une expertise dans le champ d'étude, une étude psychologique du fonctionnement de la pensée, une capacité d'observer l'élève dans l'action entreprise, un jugement équitable et respectueux de l'autre, une appropriation des techniques en évaluation.

Les formations actuelles semblent mettre l'accent sur la mécanique de l'évaluation, tels observer l'élève et s'approprier les techniques d'élaboration d'instruments. Peu de formations mettent l'accent sur les relations avec l'élève lors de situation d'évaluation. On remarque aussi qu'il existe peu de formation sur le jugement à porter et sur l'équitabilité de ce jugement. Leclerc *et al.* (1992) proposent de développer la formation au jugement dès le jeune âge et, en particulier, au niveau collégial. Leur processus de formation pourrait alors être adapté aux enseignants et enseignantes qui ont à juger, à apprécier et à évaluer. Ce processus, selon les auteurs, est de mettre l'apprenant en situation, en confrontation avec lui-même et avec les autres, afin de créer un déséquilibre entre la pensée et l'action entreprise, et arriver ainsi à rendre le jugement conscient. Cette perspective nouvelle permet alors une autre orientation dans la formation continue de l'évaluation, axée sur les acteurs en interaction constante. Elle permet d'axer la formation à l'évaluation par une entrée par les didactiques plutôt que par la spécialisation (Bain, 1988).

2.3 Projets de formation et changement planifié

Une démarche de formation continuée en enseignement et en évaluation a été développée par un groupement de conseils scolaires ontariens francophones et le ministère de l'Éducation de l'Ontario[2]. Ce programme inspiré de la démarche de changement planifé de Fullan

[2] Ce projet appelé PEPO (Projet d'excellence pédagogique de l'Ontario) a duré quatre ans et a contribué à la formation sur mesure des enseignants de sept écoles du nord est de l'Ontario, en collaboration avec les deux universités francophones et les bureaux régionaux du Ministère.

(1991) semble démontrer qu'une formation assidue et centrée sur la réalité de l'enseignement et en fonction des besoins des intervenants porte fruits. Les changements se mettent en place en intégrant les valeurs, les prises de conscience et les forces qu'une telle appropriation de nouvelles méthodologies donne aux individus impliqués. Ce programme est basé sur une démarche en quatre points : la réflexion, l'élaboration, l'expérimentation et l'objectivation. Les enseignants et les enseignantes vivent chacune de ces étapes en établissant un consensus d'équipe-école pour chacune des décisions relatives à cette formation. Le projet devient un projet école et non un projet de chef d'établissement.

Le postulat de base de tels programmes de formation continue est d'assurer un suivi constant et implique un budget et du temps. Le changement s'amorce difficilement, il est long à entreprendre et surtout il demande une constante réévaluation du vécu, de la formation et des orientations mises en place au départ. PEPO a permis à sept écoles de vivre différemment leur évolution en matière d'évaluation et d'apprentissage en assurant un support constant dans le milieu et en permettant à chacun d'évoluer à son rythme.

Les structures de formation continue, quoiqu'elles soient bien développées semblent mal définies, car elles ne mettent pas en application les principes des théories de l'apprentissage et ne permettent pas à l'apprenant/enseignant de partir de son savoir expérentiel, de l'analyser et de s'approprier une nouvelle démarche. Par ailleurs, Savoie-Zajc (1989) compare cinq modèles de gestion du changement en vue d'adopter et d'implanter une innovation, De Ketele (1988) précise les prémices d'une formation par la préparation en fonction d'une analyse des besoins auprès de la population; Artaud (1989) indique qu'il faut redéfinir pour mieux transmettre. Ainsi, la tâche des formateurs de formateurs serait de mieux définir, avec les personnes impliquées, ce qu'il faut faire pour agir avec cohérence, avec simplicité et en fonction de son intériorité.

Ce type de formation exige du formateur de développer d'abord, avec les enseignants ou enseignantes (et futurs enseignants ou enseignantes), les habiletés à réfléchir rationnellement sur leurs propres

actions d'évaluation et sur les mécanismes personnels qu'elles dégagent. Cette première étape de réflexion amènerait les intervenants et les intervenantes à situer leurs expériences de travail dans une problématique d'action humaine, complexe, qui demande une éthique professionnelle et qui exige un code moral élevé, car cette action met des êtres humains en relation soit de pouvoir, soit de collaboration. Dans un deuxième temps, il s'agirait d'élaborer des situations innovatrices d'évaluation avec les participants et les participantes, tout en tenant compte de sa marge de liberté dans le système, en sachant composer avec cette marge et surtout en sachant prendre les décisions à la fois rationnelles et subjectives qui s'imposent (Weiss, 1992). Cette formation exigerait aussi que le personnel administratif et décideur en place puisse s'assurer que l'évaluation n'est pas détachée de l'enseignement, que cet acte est subjectif, que les jugements ne peuvent pas être uniquement rationnels, scientifiques et techniquement parfaits, car l'enseignant ou l'enseignante est une personne qui a développé sa propre fiabilité dans ses jugements, à partir d'expertise, d'essais et d'erreurs et de tâtonnements. Ceci faciliterait les troisième et quatrième étapes de la formation, à savoir l'expérimentation et l'objectivation.

En formation initiale, il est judicieux de penser que la formation à l'évaluation doit suivre le même cheminement qu'en formation continue, idéalement dans un programme d'intégration de l'évaluation aux apprentissages (Bélair, 1992). Bain (1988) affirme qu'il est difficile de dissocier l'évaluation de la didactique; Allal (1988) suggère une formation transdiciplinaire où l'évaluation prend sa place à l'intérieur de ces didactiques tout en étant une discipline à part. La formation conjointe du primaire développée à l'Université d'Ottawa permet cette intégration des activités d'évaluation aux diverses activités d'apprentissage de l'étudiant-maître par l'élaboration d'un seul projet de classe intégrateur pour tous les cours de didactiques et de fondements de la formation (Bélair, 1992).

Il semble cependant que l'approche actuelle de la plupart des formations en évaluation semble occulter cette relation avec les didactiques, tout comme elle semble oublier l'habileté à communiquer qu'elle exige de l'enseignant ou de l'enseignante et de l'élève. Il serait utile de déterminer des standards communs de formation afin que le novice soit

en mesure d'évaluer, en toute connaissance de cause, l'apprentissage de ses élèves et afin que l'enseignant ou l'enseignante puisse être aussi expert dans cet aspect de son travail qu'il l'est dans l'enseignement.

CONCLUSION

Y a-t-il des solutions en vue d'une responsabilisation de l'acte d'évaluation dans l'enseignement? Les solutions possibles à la responsabilisation de l'acte d'évaluer touchent surtout la formation des personnes en action. La formation initiale devrait permettre au futur enseignant ou à la future enseignante de dégager ses forces comme évaluateur, de prendre position par rapport à certaines pratiques courantes dans le domaine et surtout d'amorcer une réflexion sur ses croyances, sur ses paradoxes et sur ses convictions profondes par rapport au système scolaire, aux échecs des élèves et à son rôle comme enseignant ou enseignante. Le décrochage scolaire augmente de plus en plus; les échecs sont de plus en plus fréquents; la motivation des jeunes est à la baisse. En Ontario, seulement 5 % de la population francophone se dirigent vers des études universitaires et près de 50 % ne terminent pas leurs études secondaires. Cette situation est alarmante et seule une prise de conscience claire et profonde de chacun des intervenants et des intervenantes dans ce système permettra de retrouver un équilibre.

La formation continue a un rôle prépondérant à remplir. Elle devrait être axée sur les actions immédiates des personnes impliquées afin de poursuivre vers une appropriation de leur système de valeur en ce qui a trait à l'enseignement et à l'évaluation des apprentissages. Des projets comme PEPO, basé sur le changement planifié, sont les moteurs de l'innovation et des changements à amorcer compte tenu des perspectives de l'an 2000. Cette formation continue devra donc «déconstruire» les principes et les règles actuelles, afin de «construire» d'autres règles et de «transformer» la relation entre l'évaluateur et l'évalué.

L'enseignement n'est plus seulement un ensemble de leçons donné à trente jeunes; c'est aussi du travail social, de l'introspection psychologique, de l'aide familiale et de l'insertion culturelle dans le milieu. Dans les années cinquante, on qualifiait l'enseignement de vocation; dans les années soixante, ce concept est devenu un métier; à l'aube de

l'an 2000, l'enseignement est considéré comme une profession. Cette profession implique un réseau de communication entre divers types de personnes, en situation plus ou moins symétrique de pouvoir, qui oblige d'établir des règles de survie en classe et en communauté scolaire. Cette profession oblige aussi à émettre des jugements qui peuvent entacher l'avenir des évalués; elle oblige à prendre des décisions malgré un certain manque d'informations qui pourraient être pertinentes et enfin, elle oblige à négocier continuellement entre la réalité de la classe, celle de l'établissement et celle de la société.

Une appropriation de l'acte d'évaluation comme un acte humain serait d'autant plus aisée que les mécanismes mis en place par le système scolaire permettront aux personnes de réfléchir sur leurs actes, sans contraintes et sans «évaluation». De plus, il faudrait que les systèmes en place soient favorables à une démarche où ce n'est plus la compétition qui est souhaitable, mais la confiance en soi et en ses capacités.

Les personnes impliquées ont des deuils à faire, des changements à apporter qui peuvent «faire mal»; les systèmes scolaires et les établissements ont aussi un rôle à jouer et des deuils à assumer. La formation initiale et continue des enseignants ou des enseignantes sur l'acte d'évaluer, l'information de plus en plus précise aux parents et l'implication de plus en plus grande de l'administration pourraient faire en sorte que la véritable fonction de l'évaluateur soit précisée et comprise par lui-même (que ce soit l'enseignant ou l'élève) et les autres intervenants.

Le rôle de la recherche serait alors de cerner le phénomène et d'apporter les solutions ou, du moins, des idées de solutions. Le rôle des systèmes scolaires serait de mettre en place des mécanismes pour que l'évolution des perceptions aille dans le sens de la véritable communication entre deux êtres humains. Le rôle des humains en action d'évaluer serait alors de rendre cette communication la plus vraie et la plus humaine possible.

RÉFÉRENCES

ALLAL, L. (1992).
Vers une pratique de l'évaluation formative. Bruxelles : De Bœck.

ALLAL, L. (1992)
Évaluation formative des processus d'apprentissage : le rôle des régulations métacognitives. Symposium de Sherbrooke sur l'évaluation des apprentissages. Sherbrooke : Université de Sherbrooke.

ALLAL, L. (1988).
Pour une formation transdisciplinaire à l'évaluation formative. *In* Gather Thurler, M. et Perrenoud, P. (éd.), *Savoir évaluer pour mieux enseigner. Quelle formation des maîtres?* (Cahier 26, 39-56). Genève : Service de la recherche sociologique.

ALLAL, L. (1988).
Vers un élargissement de la pédagogie de maîtrise : processus de régulation interactive, rétroactive et proactive. *In* Huberman, M. (éd.), *Maîtriser les processus d'apprentissage. Fondements et perspectives de la pédagogie de maîtrise* (86-126). Paris : Delachaux et Niestlé.

ANGERS, P. (1992).
La formation du jugement. *In* Schleifer, M. (éd.), *La formation du jugement : Peut-on apprendre le jugement?* (135-158). Montréal : Logiques.

ARTAUD, G. (1989).
L'intervention éducative. Ottawa : Presses de l'Université d'Ottawa.

BAIN, D. (1988).
L'évaluation formative fait fausse route : de là, la nécessité de changer de cap. *Mesure et évaluation en éducation, 10*(4), 23-32.

BÉLAIR, L. (à paraître).
La formation conjointe à l'enseignement primaire : une collaboration active avec le milieu. *Pédagogiques – Academia*

BÉLAIR, L. (1990).
Évaluation : comparer en toute équité; démarche de réflexion. *Mesure et évaluation en éducation, 13*(2), 23-46.

BLOOM, B.S., HASTING, J.T. et MADAUS, G. (1971).
Handbook on Formative and Summative Evaluation. New York : McGraw-Hill.

CARDINET, J. (1986).
Évaluation scolaire et pratique. Pédagogies en développement. Bruxelles : De Bœck.

CARDINET, J. (1986).
Évaluation scolaire et mesure. Pédagogies en développement. Bruxelles : De Bœck.

CAZABON, B. (1990).
Les habiletés d'apprentissage. Module de formation, Projet d'excellence pédagogique. Timmins : Centre de formation continue.

DASSA, C. (1992).
Régulation des pratiques, contexte de mesure et évaluation informatisée. Symposium de Sherbrooke sur l'évaluation des apprentissages. Sherbrooke : Université de Sherbrooke

DE KETELE, J. M. (1992).
Objectifs terminaux d'intégration et transfert de connaissances. Symposium de Sherbrooke sur l'évaluation des apprentissages. Sherbrooke : Université de Sherbrooke.

DE KETELE, J.M. *et al.* (1988).
Guide du formateur. Pédagogies en développement. Bruxelles : De Bœck.

DOYON, C. et JUNEAU, R. (1991).
Faire participer l'élève à l'évaluation de ses apprentissages. Montréal : Beauchemin.

FONTAINE, F. et TRAHAN, M. (1990).
Analyse du domaine des croyances reliées à l'évaluation des apprentissages. *Mesure et évaluation en éducation, 13*(1), 5-28.

FULLAN, M. (1991).
The New Meaning of Educational Change. Toronto : OISE and Teacher College Press.

HUBERMAN, M. (éd.). (1988).
Maîtriser les processus d'apprentissage. Fondements et perspectives de la pédagogie de maîtrise. Paris : Delachaux et Niestlé.

KOHLBERG, L. (1969).
Stage and Sequences : The Cognitive Developmental Approach to Socialization. *In* Goslin D. (éd). *Handbook of Socialization, Theory and Research*, (p. 348-480). Chicago : Rand McNally.

KRATHWOOL, D. (1976).
Taxonomie des objectifs pédagogiques. (Tome *2 — Le domaine affectif*). Québec : PUQ.

LECLERC, C-A. *et al.* (1992).
Intervention en contexte professionnel. *In* Schleifer, M. (éd.), *La formation du jugement : Peut-on apprendre le jugement?* (p. 239-254). Montréal : Logiques.

MINISTÈRE DE L'ÉDUCATION DU QUÉBEC. (1985).
Document d'information : *Vers un bulletin scolaire renouvelé, orientations du Ministère en notation, consignation, transmission et conservation des résultats scolaires, classes du primaire et du secondaire.* Québec : Ministère de l'Éducation.

MINISTÈRE DE L'ÉDUCATION DU QUÉBEC. (1984).
Éléments de docimologie. (Tome 1 — *La démarche évaluative*). Québec : Ministère de l'Éducation.

OUELLETTE, L-M. (1990).
La communication comme support théorique à l'évaluation formative. *Mesure et évaluation en éducation, 13*(2), 5-21.

PARÉ, A. (1977).
Créativité et pédagogie ouverte. Laval : NHP.

PERRENOUD, PH. (1992).
Du soutien pédagogique à une vraie différenciation de l'enseignement : évolution ou rupture? Tessin : Scuola Tincinese.

PERRENOUD, PH. (1991).
Différenciation de l'enseignement : résistances, deuils et paradoxes. Texte inédit. Genève : Service de recherche sociologique.

PERRENOUD, PH. (1991).
Ambiguïtés et paradoxes de la communication en classe; toute interaction ne contribue pas à la régulation des apprentissages. *In* Weiss, J. (éd.), *L'évaluation: problèmes de communication*. Suisse : ADMÉÉ-Europe, DelVal-IRDP.

PERRENOUD, PH. (1989).
L'évaluation entre hier et demain; le point de vue d'un sociologue. *Coordination, 35*, 3-5.

PERRENOUD, PH. (1988).
Nouvelles didactiques et stratégies des élèves face au travail scolaire. *In* Montandon, C. et Perrenoud, P. (éd.), *Qui maîtrise l'école? Politiques et pratiques des acteurs* (p. 175-195). Lausanne : Réalités sociales.

REBOUL, O. (1992).
Qu'est-ce que le jugement? *In* Schleifer, M. (éd.), *La formation du jugement : Peut-on apprendre le jugement?* (p. 17-36). Montréal : Logiques.

ROBILLARD, R. (1992).
La place primordiale à la personne de l'éducateur pour le changement de sa pratique évaluative. Symposium de Sherbrooke sur l'évaluation des apprentissages, Université de Sherbrooke.

ROBILLARD, R. (1990).
Évaluer, un acte d'intériorité. Conférence donnée à la 13e session d'études de l'ADMÉÉ.

SAVOIE-ZAJC, L. (1990).
Un modèle de changement planifié. *In* Actes du colloque de l'AIPELF : *Les modèles en éducation*.

SAVOIE-ZAJC, L. (1989).
Cadres conceptuels et évolution des modèles de changement planifié en éducation. *Revue des sciences de l'éducation, 15*(1), 123-138.

SCALLON, G. (1988).
L'évaluation formative. (Tome I — *La réflexion*). Québec : Université Laval.

SIROTA, R. (1988).
L'école primaire au quotidien. Paris : PUF.

TARDIF, J. (1992).
L'évaluation dans le paradigme constructiviste. Symposium de Sherbrooke sur l'évaluation des apprentissages, Université de Sherbrooke.

VAN NIEUWENHOVAN, C. (1990).
Analyse des représentations des enseignants primaires en Belgique. Mémoire de licence. Belgique : Université Louvain-La-Neuve.

WEISS, J. (1992).
L'enseignant «au cœur froid» ou l'objectivité en évaluation». *Mesure et évaluation en éducation, 14*(4), 19-32.

LA PLACE PRIMORDIALE À LA PERSONNE DE L'ÉDUCATEUR POUR LE CHANGEMENT DE SA PRATIQUE ÉVALUATIVE

RICHARD ROBILLARD
Université Sainte-Anne

INTRODUCTION

Il y a maintenant plus de dix ans que je suis engagé dans l'étude sur la nature du phénomène de l'évaluation en éducation et également sur la pratique évaluative de l'éducateur. Je cherche en effet à comprendre et à éclairer, pour moi et pour d'autres, le sens et le rôle fondamental de l'évaluation en éducation d'une part, et à parvenir à ce que ce puissant et complexe acte humain soit effectué et vécu dans un cadre de respect et d'aide au développement de chaque acteur (éducateur comme élève), d'autre part. Autrement dit, faire de l'évaluation un acte humain, développemental et transpersonnel. En fait, je cherche à faire de l'évaluation un acte d'éducation, avec tout ce que cela peut exiger, sans pour autant négliger son importance pour la réalisation des apprentissages, entre autres, ceux qui sont prédéterminés et dictés dans les programmes d'étude.

Je vois l'évaluation comme un acte qui permet à la personne évaluée de se sentir respectée dans ce qu'elle est comme apprenante, ce qui n'élimine pas la confrontation souvent inhérente à cet acte; et ensuite d'être aidée à prendre conscience de sa situation et de l'évaluation de ses apprentissages ainsi que des changements possibles à réaliser et de sa capacité à y parvenir d'une façon personnelle. Alors tout acte d'évaluation ne se limite pas aux apprentissages. C'est un acte qui touche la personne dans son entité et qui a une résonance jusqu'au plus profond d'elle-même.

Précisons un peu plus! Je sens le besoin d'apporter quelques précisions sur ce qui m'a amené à cette vision et à cette position de l'évaluation en éducation. Faire une place à la personne de l'éducateur, c'est lui permettre de clarifier et d'exprimer sa perception de l'évaluation. D'abord, cette orientation de recherche et ce choix pour une pratique évaluative qui soit, d'abord et avant tout, éducative sont soutenus par

mon histoire de vie ainsi que par mon vécu quotidien par rapport à l'évaluation. Je réalise aujourd'hui que je me débats encore dans ma vie personnelle, d'éducateur et de formateur avec les empreintes laissées par les expériences intérieures vécues autrefois, durant l'enfance et l'adolescence, lors de situations formelles et informelles d'évaluation, ce que j'appelle des «expériences d'évaluation». Alors, en utilisant différentes approches qui mettent l'accent sur la découverte et sur la compréhension du monde intérieur de la personne, je travaille à relativiser, à digérer et à intégrer ces expériences d'évaluation éprouvées à la suite des actes évaluatifs que des adultes ont posés explicitement et implicitement, consciemment et inconsciemment, à mon égard ou sur d'autres élèves. Ce travail personnel d'intégration de ces expériences d'évaluation m'apparaît essentiel à réaliser, car je suis maintenant persuadé qu'elles sont à l'origine du développement de la perception de l'évaluation ainsi que de la pratique évaluative que je vivais lorsque j'ai débuté comme enseignant; cette perception et cette pratique se sont transformées depuis que je cherche à comprendre le sens de l'évaluation, à intégrer ces expériences d'évaluation et à m'approprier ma pratique évaluative, trois gestes qui ne peuvent venir que de l'intérieur de la personne et qui ne peuvent jamais être imposés de l'extérieur.

Ce retour et ce travail sur mes expériences d'évaluation éprouvées autrefois a pour effet, entre autres, de prendre conscience de toute la puissance et de la complexité de l'acte d'évaluer des apprentissages, de son impact sur le développement de la personne de même que de la responsabilité dont chaque éducateur, auteur de son acte d'évaluation, doit récupérer et s'approprier dans l'exécution de sa pratique évaluative.

Il m'apparaît plausible que ce rapport entre le pouvoir de l'évaluation et le développement de l'apprenant ait pu influencer et guider, du moins inconsciemment, quelques-unes des recherches, des réflexions et des théories sur l'acte d'évaluer des apprentissages. C'est ce que je perçois dans les efforts investis depuis les dernières années pour l'application de l'évaluation formative, une évaluation qui se préoccupe davantage de la personne de l'évalué apprenant à cause, justement, de la fonction «régulatrice» de ses apprentissages (Allal, Cardinet et Perrenoud, 1979; Dassa, 1989; Perrenoud, 1991; Scallon, 1988a, 1988b).

Il ne fait aucun doute que l'évaluation formative est un acte essentiel pour tout apprenant. Mais cet acte me semble incomplet, car c'est presque uniquement la «maîtrise des apprentissages prédéterminés» qui est visée et non pas le développement intégral de la personne. L'évaluation formative, telle qu'elle est prescrite et appliquée, se limite surtout à l'atteinte des objectifs présélectionnés; elle n'offre pas à la personne de l'évalué la possibilité de se connaître davantage et d'apprendre graduellement à prendre en main ses propres apprentissages : le but ultime de l'éducation.

> *L'être humain est formé de plusieurs entités : moral, physique et spirituel qui se révèlent par ses sentiments, ses concepts, ses paroles, ses attitudes corporelles, son état conscient et même par son subconscient. Un des buts de l'éducation est d'assurer une harmonieuse intégration de ces entités, en se connaissant, au point même de débusquer les motifs inconsciemment ou consciemment cachés de notre agir* (Desjardins, 1992, 110).

L'acte d'évaluer des apprentissages, que ce soit avec ou sans la fonction régulatrice, devient éducatif, donc un geste de formation, lorsqu'il développe, chez l'évalué, cette capacité à donner du sens à ce qu'il vit intérieurement en rapport avec ce qu'il manifeste comme apprentissage. Cette formation dans l'acte d'évaluer des apprentissages se réalise par ce «souci de liaison constante entre d'un côté l'hétéroformation et de l'autre, l'autoformation : c'est-à-dire entre une formation où prédomine l'intervention d'agents extérieurs et celle qui repose essentiellement sur une démarche personnelle, significative et autonome de la personne elle-même» (Gingras, 1986).

Cela m'amène à une autre précision quant à l'orientation de mes recherches et réflexions sur l'acte d'évaluer des apprentissages. Il s'agit de l'éducation à l'évaluation. C'est ainsi que je conçois mon travail d'éducateur et de formateur. À travers ma pratique éducative et évaluative, favoriser, entre autres, chez les futurs enseignants, non seulement l'apprentissage d'un acte d'évaluation qui respecte et aide l'évalué dans son développement; mais également leur permettre d'apprivoiser l'évaluation, apprendre les principes, règles et limites de fonctionnement de cet acte humain, découvrir et intégrer les différents processus et minipro-

cessus inhérents à l'acte d'évaluer, saisir la pertinence, le sens et l'impact de cet acte sur le développement interne de l'élève, en particulier au niveau de la conscience de soi et de son centre interne d'évaluation, deux facultés primordiales pour le développement de la personne; et enfin, la recherche, l'appropriation et l'actualisation d'une pratique évaluative qui soit éducative.

Cette éducation à l'évaluation m'apparaît une nécessité à favoriser à travers nos actes d'évaluation des apprentissages; et ce, à tous les niveaux scolaires. Cette éducation à l'évaluation dont je n'ai pas reçue durant l'enfance m'a manqué tout au long de ma formation. Il faut également mentionné que cette éducation à l'évaluation passe par la pratique évaluative et éducative de l'éducateur, c'est-à-dire par l'ensemble de ses actes évaluatifs constamment effectués. C'est donc par sa pratique évaluative que l'éducateur favorise ou non l'éducation à l'évaluation, en particulier chez l'enfant et chez l'adolescent. Cela fait voir la puissance que possède chacun des actes d'évaluation de l'éducateur.

N'est-ce pas, peut-être, à cette préoccupation d'éducation à l'évaluation que voulait nous sensibiliser et nous confronter Pourtois (1979) par sa question : «Quel profil d'homme se dévoile à travers l'actuelle technologie de l'évaluation?» N'est-ce pas également vers cette voie éducative que s'orienteraient les travaux de l'équipe d'Aix sur ce que Bonniol et Nunziati ont appelé l'évaluation formatrice (Perrenoud, 1991, 61)? Enfin, n'est-ce pas également que cet autre but, trop souvent délaissé, à mon avis, de l'évaluation formative, peut-être parce qu'on ne sait pas encore comment l'intégrer à notre pratique évaluative, soit l'aide au développement de la personne? «Est formative toute évaluation qui aide l'élève à apprendre et à se développer, autrement dit, qui participe à la régulation des apprentissages et du développement dans le sens d'un projet éducatif» (Perrenoud, 1991, 50). Dans ce cadre, ce n'est donc plus uniquement une évaluation formative, mais également une évaluation formatrice et éducative.

S'il est important de «savoir comment l'évaluation formative aide l'élève à apprendre, par quelles médiations elle rétroagit sur les processus d'apprentissage», comme le mentionne Perrenoud (1991, 50), il importe également de savoir comment l'acte d'évaluer des apprentissages

contribue au développement de la personne de l'évalué ainsi qu'à son éducation à l'évaluation. C'est avec l'intériorisation de ses expériences d'évaluation, du développement de sa pensée évaluative et de son éducation à l'évaluation que tout individu doit nécessairement composer durant son existence. En fait, il ne s'agit pas seulement de former l'élève à la régulation de «ses propres processus de pensée et d'apprentissage», ce que préconise l'évaluation formative (Perrenoud, 1991), mais surtout de l'aider à ce qu'il se crée une perception et une pratique d'évaluation qui soit éducative, autant pour lui-même que pour son entourage immédiat et futur.

C'est cette option d'éducation à l'évaluation qu'il m'apparaît souhaitable d'explorer et de vivre dans les prochaines années. Une option qui fait suite à ce qui a été réalisé jusqu'à présent en évaluation des apprentissages, soit l'intensification d'une pratique régulatrice des apprentissages qui est venue bousculée la pratique davantage comparative et normative des apprentissages.

Abordons maintenant le centre de l'acte d'évaluer des apprentissages. Il ne fait plus aucun doute que l'acte d'évaluation ne se limite pas uniquement aux apprentissages. C'est un acte qui atteint la personne de l'évalué dans le sens qu'il engendre certaines expériences, affecte son développement et influence son éducation à l'évaluation. Cette réalité doit être prise en considération même si, pour l'instant, nous marchons davantage dans l'obscurité et l'inconnu que dans la clarté.

Un regard plus approfondi de cette réalité permet de relever la constatation suivante : ce qui atteint la personne de l'évalué, ce sont les actes évaluatifs effectués par l'éducateur, autrement dit sa pratique évaluative. Nous voilà alors au centre de l'acte d'évaluer des apprentissages, c'est-à-dire là où réside la pratique évaluative de l'éducateur.

Cette constatation concernant le rapport entre la personne de l'évalué et la pratique évaluative de l'éducateur peut s'avérer quelque peu dérangeante puisqu'on s'adresse directement à l'éducateur et on questionne la nature et l'impact de ses actes d'évaluation. Il s'agit, certes, d'un discours différent de ce qu'on entend habituellement, car on ne parle pas de l'acte d'évaluation en général, mais spécifiquement de son auteur.

Évidemment, il n'est nullement question de juger et encore moins de blâmer l'éducateur ou sa pratique évaluative, gestes tout à fait impensables et surtout infaisables lorsqu'on connaît et comprend la structure de la pratique de tout éducateur. La pratique évaluative de chacun d'entre nous est constituée de réactions et de motifs internes qui nous échappent le plus souvent qu'autrement. Alors notre responsabilité est justement de revenir sur ses propres gestes évaluatifs, de reprendre contact avec certains moments d'intervention afin d'y déceler et de comprendre les motivations profondes qui agissaient inconsciemment sur sa pratique évaluative. «Ces prises de conscience sont vraisemblablement ce qui peut le mieux influencer les actions à venir» (Paré et Auclair, 1988, 31). Ce travail d'introspection ne peut se faire que par et sur soi-même, ni être imposé de l'extérieur.

De plus, tout éducateur ignore la réelle portée de sa pratique évaluative sur la personne de l'évalué. Encore ici, il y a trop d'éléments qui échappent à l'éducateur et qui n'appartiennent qu'à l'évalué. Toutefois, cela ne justifie nullement que l'on ne se préoccupe pas de la personne de l'évalué. Cette limite nous ramène simplement à notre responsabilité et au pouvoir que nous possédons de travailler à devenir davantage conscient de sa propre pratique évaluative afin qu'elle puisse, le mieux possible, respecter la personne de l'évalué et l'aider dans son développement et son éducation à l'évaluation.

Ce qui se révèle intéressant dans ce rapport entre la personne de l'évalué et la pratique évaluative de l'éducateur, c'est la nécessité de prendre davantage en considération la pratique évaluative réelle de l'éducateur et non plus la pratique imaginée et planifiée par des agents externes. On peut, par exemple, commencer à se demander ce qu'il en est vraiment de la pratique évaluative de l'éducateur. Cette question nous ramène nécessairement à interroger sa propre pratique. Quelle est la nature de sa propre pratique évaluative? Comment sa propre pratique évaluative peut-elle affecter la personne de l'évalué? Qu'est-ce qui se cache derrière sa pratique évaluative? Quels sont les motifs inconscients qui agissent sur ses propres actions évaluatives? Enfin, quel changement il nous apparaît important et possible d'apporter dans sa pratique évaluative?

Faire une place à la personne de l'éducateur. Entrer en contact avec la pratique évaluative de l'éducateur, c'est nécessairement entrer en contact avec la personne de l'éducateur. Qu'en est-il de cette personne qui accomplit quotidiennement un nombre incalculable d'actes évaluatifs formels et informels, explicites et implicites, conscients et inconscients? Cette question nous interroge sur la place qui a été faite à la personne de l'éducateur dans la recherche et à la réflexion sur l'acte d'évaluer des apprentissages, d'une part, ainsi que sur l'insistance à vouloir changer sa façon d'évaluer, d'autre part.

Jusqu'à présent, on s'est surtout intéressé à la personne de l'évalué, en particulier en ce qui regarde la maîtrise des apprentissages souhaités. Cet intérêt a influencé la recherche de théories, d'instruments et de démarches d'évaluation afin de favoriser, le plus possible, l'atteinte de cette maîtrise chez l'apprenant.

Le temps est maintenant venu, sans pour autant délaisser la personne de l'évalué, de s'intéresser à la personne de l'éducateur. Ce texte tente donc de proposer une ouverture à la personne de l'éducateur dans l'acte d'évaluer des apprentissages et particulièrement en ce qui concerne le changement de sa pratique évaluative. Mon ambition est surtout d'alimenter, voire de provoquer, la réflexion sur un sujet, comme le dit Abraham (1984), «qui n'a pas encore trouvé sa place véritable dans les sciences de l'éducation», ni dans le domaine de l'évaluation des apprentissages. «La recherche dans ce domaine est pauvre, les résistances devant la découverte des faits sont énormes et les théories sont loin d'être élaborées» (Abraham, 1984, 11).

Bien que nous soyons novices dans nos réflexions et dans nos recherches sur la personne de l'éducateur dans le domaine de l'évaluation des apprentissages, il m'apparaît primordial d'y investir temps et efforts. Ce travail s'oriente dans le même sens de ce que recommandent de plus en plus de chercheurs en éducation (Abraham, 1972, 1989; Abraham *et al.*, 1984; Artaud, 1982, 1985, 1989; Bédard, 1981; 1983; 1984; Craig, 1978; Desjardins, 1992; Ditisheim, 1991; Gingras, 1986; Haramein, 1978, 1989, 1991; Huberman, 1984, 1989; Meirieu, 1991; Moyne, Artaud et Barlow, 1988; Paré, 1983, 1989; Paré et Auclair, 1988; Vassileff, 1992) : se centrer sur la personne de l'éducateur, qu'il soit son

point de référence, qu'il soit autonome et lui-même professionnellement, qu'il travaille à sa propre création, qu'il se réapproprie sa formation de même que sa pratique éducative, qu'il se connaisse davantage, qu'il devienne conscience du rapport entre sa transformation intérieure et le changement de sa pratique éducative.

Dans ce qui suit, je me propose de regarder a) une rétrospective de mes expériences d'intervenant en évaluation; b) le passage de la recherche d'un acte «idéal» d'évaluation à la pratique évaluative «réelle» de l'éducateur; et c) trois questions sur la place à accorder à la personne de l'éducateur pour le changement de sa pratique évaluative : Que signifie «faire une place à la personne de l'éducateur? Qu'est-ce qui nous motive à faire une place à la personne de l'éducateur? Comment faire une place à la personne de l'éducateur pour le changement de sa pratique évaluative?

1. RÉTROSPECTIVE DE MES EXPÉRIENCES D'INTERVENANT EN ÉVALUATION

Avant d'aborder la problématique en question, il me semble important de dresser une brève rétrospective de mes expériences d'intervenant en évaluation, parce que ces expériences ont influencé l'évolution de ma perception de l'évaluation, l'orientation de mes recherches et de mes réflexions au sujet de l'évaluation des apprentissages et surtout ma conviction profonde de la nécessité de faire une place primordiale à la personne de l'éducateur pour le changement de sa pratique évaluative.

Ce «regard en arrière» qu'offre la rétrospective est un phénomène qui permet de donner du sens à ses expériences et de s'approprier le changement de sa pratique, la responsabilité ultime de tout éducateur. «La rétrospective est un processus de revue et d'évaluation, un regard sur ce qui a été, afin de trouver un sens à la vie dans le présent pour nous permettre de faire quelque chose de mieux dans le futur» (Levin, 1992).

Ce regard dans le passé que permet la rétrospective (ou l'histoire de vie) est un geste de plus en plus préconisé en éducation, en particulier, dans la formation des enseignants (Boudana, 1992; Ditisheim, 1991; Gingras, 1986; Paré, 1991; Progoff, 1984; Vassileff, 1992). C'est un acte qui fait une place à la personne de l'éducateur.

Dans la rétrospective de mes expériences d'intervenant en évaluation, je m'attarderai à la signification de mes observations, perceptions, de mes impressions et de mes constatations; des outils intérieurs et subjectifs qui possèdent une richesse pour la compréhension de la réalité. D'autant plus que ce sont des outils humains qui sont constamment sollicités dans la pratique évaluative de tout éducateur (Robillard, 1990).

Au cours des dix dernières années, j'ai eu le privilège d'assumer diverses fonctions d'intervenant qui m'ont permis d'être en contact autant avec les acteurs impliqués dans des situations d'évaluation (évaluateur et évalué) qu'avec le phénomène de l'évaluation. Parmi ces fonctions, je retiens particulièrement celle de conseiller pédagogique en mesure et évaluation et de directeur-adjoint d'une école primaire. Dans chacune de ces deux fonctions d'intervenant, la principale tâche consistait à aider des enseignants dans leur acte d'évaluer des apprentissages, plus spécifiquement à les soutenir dans l'application des démarches évaluatives prônées, entre autres, dans les programmes d'étude.

Ce qui m'impressionnait et me questionnait dans ce travail d'aide, c'était le rapport que des enseignants entretenaient avec le phénomène de l'évaluation. Chez plusieurs enseignants, je percevais l'existence d'un malaise profond lorsqu'ils décrivaient leur façon d'évaluer les apprentissages des élèves. J'observais comment ce discours sur l'évaluation des apprentissages et surtout de ses propres actes d'évaluation était chargé de sentiments, d'expériences internes non intégrées par rapport à l'évaluation. Un discours, comme le précise Perrenoud (1984), «chargé de valeurs, qu'il soit autocritique ou autojustificateur, cynique ou idéaliste» (p. 29). Ce dont je prenais conscience, c'est que parler de sa pratique évaluative, entre autres, de sa façon d'évaluer formellement des apprentissages, est un geste où l'éducateur se rend vulnérable; c'est à la fois exposer une partie de soi-même et ouvrir la porte à son monde intérieur là où résident les raisons profondes, souvent inconscientes, qui affectent sa gestuelle évaluative. C'est d'ailleurs ce que relèvent Paré et Auclair (1988) au sujet du geste pédagogique :

> *Dans chaque geste pédagogique, il y a la personne qui intervient. Entrer en contact avec le geste pédagogique, c'est entrer*

> *en contact avec une partie de la personne qui l'a posé. Nous ne croyons pas que le geste pédagogique corresponde à la totalité de la personne. La personne ne peut jamais être réduite au geste qu'elle vient de poser. Nous croyons cependant qu'il y a dans ce geste une manifestation précise d'un aspect de la personne que nous pouvons connaître à travers ce geste et que la personne peut elle-même connaître* (Paré et Auclair, 1988, 31).

Chez certains enseignants, leur malaise se faisait davantage observable et sentir lorsqu'ils entrevoyaient la nécessité d'effectuer un changement quelconque dans leur façon d'évaluer des apprentissages. À travers leurs paroles ou leur silence, il y avait des objections et des résistances. Je réalisais que changer, si peu soit-il, son acte d'évaluation (par exemple, mettre plus d'emphase sur la régulation des apprentissages) peut s'avérer un geste difficile, voire menaçant, pour certains enseignants, surtout lorsque ce changement n'émerge pas clairement de l'intérieur de la personne, mais qu'il est perçu davantage comme imposé de l'extérieur. Il aurait été intéressant d'approfondir l'expérience intérieure que pouvaient vivre ces enseignants par rapport au changement de leur pratique évaluative; qu'est-ce qui provoquait ce malaise ou cette tension.

J'étais intrigué par le type de changement que semblaient apporter certains enseignants quant à leur acte d'évaluation. Un changement qui me laissait l'impression que ces enseignants se soumettaient aux demandes externes plutôt que de se référer à leur propre expertise et d'en prendre possession. Chez d'autres enseignants, j'avais l'impression que leur changement était effectué dans l'espoir de se débarrasser de ce «mal» que leur apparaissait l'évaluation des apprentissages; ils traduisaient verbalement cet espoir par «si je pouvais seulement enseigner et ne plus avoir à évaluer». Intérieurement, je réagissais à leur comportement, car il me semblait que ces enseignants se déconnectaient non seulement de la personne de l'élève, mais également de leur propre personne.

Je prenais conscience que, dans le domaine de l'évaluation des apprentissages, tout comme en éducation, on ne peut imposer de l'extérieur un changement réel, intégré et personnel par rapport à la pratique de l'éducateur. Le changement véritable demande du temps d'intégration,

des prises de conscience et surtout une référence à soi-même. Le changement de sa pratique évaluative peut être nourri par l'extérieur, mais il ne se commande pas; il émerge avec un travail personnel.

En regard du changement planifié avec lequel étaient confrontés des enseignants au sujet de leur acte d'évaluation, j'éprouvais un état de tension. J'étais en conflit avec l'insistance, auprès des enseignants, afin qu'ils modifient leur manière d'évaluer des apprentissages et qu'ils adhèrent aux démarches prônées dans les programmes d'étude, sans au préalable se centrer ni s'intéresser à la personne de l'éducateur. De plus, je réalisais qu'on ne peut me changer de l'extérieur, qu'aucun individu n'a le pouvoir de me changer; moi seul, je peux me transformer et changer ma pratique. Ce retour à ma situation m'aidait à comprendre ce que pouvaient éprouver certains enseignants au sujet du changement sollicité de leur acte d'évaluer des apprentissages. Je rejoins Paré (1989) quand il dit : «Il n'est pas possible de changer qui que ce soit de l'extérieur. Je n'ai pas le pouvoir de changer l'autre ni par coercition, ni par manipulation». C'est la personne, elle-même, qui décide d'être touchée et d'apporter un changement conscient à sa pratique. En fait, ce qui m'intéressait comme intervenant en évaluation, ce n'était pas tant comment l'enseignant s'y prenait pour évaluer des apprentissages, ni les démarches évaluatives préfabriquées, mais plutôt ce qui se passait à l'intérieur de cet éducateur, là où se trouvait la source de sa pratique évaluative.

J'observais, chez plusieurs enseignants, la présence d'un écart, voire d'un fossé, entre leur pratique évaluative et leur pratique pédagogique, en particulier lors des situations formelles d'évaluation des apprentissages. Dans ces moments, les activités d'évaluation différaient grandement des activités d'apprentissage; elles étaient davantage présentées d'une façon plus tendue, impersonnelle, fermée, sans mise en situation ni objectivation, ce que l'élève retrouvait souvent dans les activités d'apprentissage. C'était comme s'il passait, soudainement et probablement inconsciemment, à un autre registre gestuel et d'intervention lorsqu'il se retrouvait dans une situation d'évaluation des apprentissages. C'était un registre dans lequel il m'était possible de percevoir des gestes davantage mécanisés, techniques, imprécis, saccadés, distants, impersonnels et parfois maladroits, impatients et tendus. C'était des gestes

qui me semblaient cacher un sentiment d'inconfort, d'impuissance et de déplaisir. Par ailleurs, dans l'animation des activités d'apprentissage, les gestes de l'enseignant m'apparaissaient plus personnels, chaleureux, aidants : ils laissaient entrevoir un sentiment d'aisance et de plaisir d'être avec l'élève. Se pourrait-il que ces enseignants s'étaient appropriées leur pratique pédagogique et non pas leur pratique évaluative?

Il me semblait exister, chez plusieurs enseignants, une méconnaissance du sens et du rôle éducatif que peut avoir l'évaluation des apprentissages chez la personne de l'évalué. Cette ignorance pourrait-elle expliquer, en partie du moins, la recherche incessante et souvent insatisfaisante de méthodes, de recettes ou de trucs pour l'exécution de leur acte d'évaluation, une exécution qui devient dictée davantage de l'extérieur que de l'intérieur? L'appropriation de sa pratique évaluative ainsi que du changement de cette pratique ne repose-t-elle pas sur l'élaboration, sur la précision et sur l'expression de sa pensée évaluative, du sens et du rôle que possède, à ses propres yeux, l'évaluation des apprentissages?

Plusieurs enseignants manifestaient une certaine insensibilité et une certaine inconscience quant à l'impact que pouvait avoir leur pratique évaluative sur la personne de l'élève. Cette insensibilité me dérangeait en tant qu'intervenant surtout qu'elle réveillait des vieilles expériences d'évaluation vécues durant l'enfance et l'adolescence. Quoi qu'il en soit, cette inconscience par rapport à la puissance et aux effets possibles de ses actes d'évaluation n'est-elle pas une indication d'une certaine déconnexion avec sa pratique évaluative? Pour atteindre ce niveau de conscience ne faut-il pas être en contact avec soi-même, entre autres, avec son histoire, avec ses expériences, avec sa perception et avec sa pensée évaluative, sa dimension éthique d'éducateur? Ne faut-il pas également que la personne de l'éducateur ait une place dans la recherche, dans la réflexion et dans les changements proposés dans le domaine de l'évaluation des apprentissages?

Ce travail d'intervenant auprès des acteurs aux prises avec l'évaluation m'a amené à prendre conscience de la manière dont la pratique évaluative de l'éducateur possède un langage. Sa pratique évaluative

transmet constamment des messages évaluatifs, que ce soit par la parole, par le silence, par le regard, par les attitudes, par les comportements, par les mots ou par les chiffres écrits... De plus, la pratique évaluative de l'éducateur, tout comme sa pratique éducative, exprime et communique, à l'autre et à soi-même, des informations sur son monde intérieur, entre autres, sur la nature de ses expériences d'évaluation, sur sa perception et sur sa pensée de l'évaluation, sur ses sentiments, sur ses réflexions, sur ses doutes... La présence et la puissance de ce langage expliquent, selon moi, pourquoi la personne de l'évalué peut être tant touchée et affectée par les actes d'évaluation de l'éducateur. Une étude de ce langage s'avère alors intéressante.

Jusqu'à présent, il a été fait mention de la pratique évaluative d'enseignants en fonction. J'aimerais maintenant m'attarder à un autre type de pratique évaluative, celle que j'appelle «en herbe» et qu'on rencontre chez les enseignants en formation. Comme formateur de ces futurs formateurs, il m'a été possible de constater à quel point plusieurs d'entre eux étaient préoccupés, voire craintifs, à l'idée de devoir éventuellement poser des actes d'évaluation. Il était intéressant et également décevant de constater de quelle façon ils percevaient l'évaluation : comme un acte davantage menaçant et pénalisant plutôt que comme un geste de respect et d'aide au développement de la personne de l'évalué. Derrière cette perception de l'évaluation, il y avait des expériences davantage désagréables qu'agréables (allant jusqu'à des blessures) non encore intégrées. Ces individus étaient beaucoup plus inquiets de leur pratique évaluative, sachant son impact sur la personne de l'évalué, que de leur pratique pédagogique. Il m'apparaît alors important de regarder comment il serait possible d'aider ces futurs enseignants, durant leur période de formation, à commencer l'appropriation de leur pratique évaluative afin de la rendre éducative. Il ne fait aucun doute que ce début d'appropriation ne peut se réaliser qu'en faisant une place à la personne de ces futurs éducateurs.

Cette rétrospective sur mes expériences d'intervenant en évaluation me permet de me rendre compte de la manière dont les informations empiriques obtenues d'éducateurs en fonction et en formation ont été révélatrices et soutenantes, non seulement pour l'orientation de mes re-

cherches en évaluation et ma conviction de la place que doit maintenant prendre la personne de l'éducateur, mais également pour la place que je dois davantage accorder à ma propre personne dans mon travail d'appropriation constante de ma pratique évaluative. Il m'apparaît essentiel de m'impliquer en tant que chercheur et de vérifier, sur moi-même, mes réflexions et mes conclusions. À quoi servirait, en effet, un discours théorique sur la place que doit prendre la personne de l'éducateur dans le domaine de l'évaluation des apprentissages et le changement de son acte d'évaluation si, quotidiennement, je ne vis pas l'expérience de faire une place prioritaire à ma propre personne. Comme Vassileff (1992), je tiens à rompre «avec cette tradition qui voudrait que, pour étayer une théorie, on l'assortisse d'observations dont serait exclu l'observateur. Je crois, au contraire, qu'un chercheur doit s'impliquer personnellement, de la manière la plus transparente possible, pour tendre vers une «science avec conscience» (p. 14).

Il y a de plus en plus de chercheurs (Craig, 1978; Haramein, 1991; Moustakas, 1990; Vassileff, 1992) qui s'inscrivent dans cette philosophie de recherche en sciences de l'éducation où ils tiennent compte également de la richesse des expériences du chercheur et qu'ils désirent que la recherche devienne un lieu de formation pour le chercheur et qu'elle lui procure une compréhension plus approfondie de la structure et de la dynamique de sa propre pratique éducative, de son monde intérieur et du rapport entre ces deux entités.

2. De la recherche d'un acte idéal d'évaluation des apprentissages à la pratique évaluative «réelle» de l'éducateur

Cette rétrospective personnelle qui vient d'être faite nous permet de percevoir deux constatations importantes. D'abord, de réaliser à quel point la personne de l'éducateur est le fondement de l'acte d'évaluer des apprentissages. Ensuite, de prendre conscience de la façon dont l'étude de la pratique réelle de l'éducateur ainsi que de son monde intérieur et de l'interaction entre ces deux entités n'a pas été, au cours des deux dernières décennies, au centre des préoccupations dans nos recherches et dans nos réflexions en mesure et évaluation.

Ces deux constations, ou plutôt ces deux réalités, peuvent, à première vue, être dérangeantes et paraître quelque peu insignifiantes, surtout lorsqu'on s'attarde à regarder du côté de l'évaluation des apprentissages prédéterminés que l'élève doit maîtriser. Ces constatations, même si elles demeurent empiriques, doivent, à mon avis, être simplement accueillies dans l'unique but de poursuivre et de nourrir l'évolution de nos recherches et de nos réflexions dans le domaine de l'évaluation des apprentissages.

2.1 La recherche d'un acte idéal d'évaluation des apprentissages

Depuis vingt ans, des efforts considérables ont été déployés pour renouveler, pour modifier, voire pour améliorer, l'acte d'évaluer des apprentissages des enseignants. Cette entreprise a donné lieu à des recherches, à des réflexions et à des théories dans les différents secteurs reliés à l'évaluation des apprentissages. Il y a eu, par exemple, une recherche de qualité et de pertinence dans la construction des instruments pour recueillir des informations quant aux apprentissages des élèves. Il en a été de même en ce qui concerne l'analyse des données recueillies, la planification des objectifs d'évaluation en rapport avec ceux de l'enseignement, la communication des résultats. Quoi dire de l'intérêt manifesté pour l'application et pour la praticabilité de l'évaluation formative. Il ne faut surtout pas omettre de mentionner la quantité des séances de perfectionnement distribuées aux enseignants sur l'acte d'évaluer des apprentissages en fonction de la démarche évaluative présentée et prônée dans les divers programmes d'étude.

Tous ces efforts, ceux d'hier de même que ceux d'aujourd'hui, ne s'orientaient-ils pas, du moins implicitement, vers la production, chez les enseignants, d'un acte idéal d'évaluation des apprentissages? Idéal dans le sens que leur acte d'évaluation rejoint, le plus possible, les deux attentes suivantes : appliquer la «régulation des apprentissages» et parvenir à l'attribution d'un jugement de valeur qui soit équitable, fidèle et précis. C'est sur ces deux attentes sociales que se sont greffées de nombreuses recherche et théories, parfois «magiques» (Bélair, 1991) dans le domaine de l'évaluation des apprentissages.

Je m'empresse de mentionner que la recherche de l'idéalisme, autant dans l'acte d'évaluer des apprentissages que dans les autres gestes éducatifs, n'est pas à rejeter. D'ailleurs, quel intervenant préoccupé par la qualité de l'évaluation n'a-t-il pas, ne serait-ce qu'une seule fois, imaginé ou rêvé d'atteindre la perfection dans l'acte d'évaluer des apprentissages? Je sais que c'est cette recherche inconsciente d'un acte idéal d'évaluation qui influençait mes interventions auprès des enseignants et qui est encore présente dans mon travail de formateur de futurs enseignants : l'espoir, et parfois le rêve, de les former parfaitement, sur tous les plans (théorique, technique, pratique) concernant leur pratique évaluative ainsi que l'espoir de développer chez ces derniers la sensibilité et l'habilité à respecter l'évolution de chaque élève et à l'aider à prendre conscience de l'état de ses apprentissages, de favoriser la régulation des apprentissages, de permettre, chez chacun des élèves, une éducation à l'évaluation, d'être conscient de l'impact possible de sa pratique évaluative sur le développement de l'élève, de situer, le plus exactement possible, les apprentissages réalisés par l'élève tout en évitant d'émettre des «jugements de valeur» à son égard, enfin, d'utiliser des instruments appropriés pour recueillir, pour analyser et pour communiquer des informations pertinentes quant à l'état des apprentissages de l'élève.

Artaud (1989) décrit fort bien la structure de cet idéalisme qui a orienté nos recherches, nos théories et nos pratiques, autant dans le domaine de l'évaluation des apprentissages que dans les modèles d'intervention éducatifs :

> *L'adulte qui s'engage dans une démarche éducative se propose en effet un but à atteindre qui est indissociable de l'image idéale de lui-même qu'il s'est forgée au cours de son développement et qui constitue le mobile de ses actions. Ce modèle, même s'il n'est pas clairement explicité, ne cesse de se manifester à travers ses manières d'agir et de réagir dans les situations éducatives, à travers la conception qu'il se fait de son rôle, le regard qu'il porte sur l'enfant, la réussite sociale qu'il rêve pour lui et les valeurs qu'il veut lui transmettre. Chaque fois qu'il intervient pour encourager des comportements jugés souhaitables, corriger ou condamner des comportements considérés comme déficients, il a recours à des critères d'évaluation qui reflètent la conception*

> *qu'il se fait du fonctionnement optimal d'une personnalité. Il en est de même pour l'institution éducative qui, de concert avec la famille, prend en charge l'éducation de l'enfant : sa structure son mode de fonctionnement, ses procédures éducatives sont étroitement dépendants du modèle humain qu'elle entend promouvoir. Quand elle fixe dans ses programmes les contenus de connaissances à assimiler, les modalités d'apprentissage et les exigences d'évaluation, quand elle élabore des règlements qui décrivent les conduites souhaitables et le style d'interaction que l'enseignant doit avoir avec ses élèves, elle privilégie un type de personnalité et de rapport sociaux bien définis. Là encore, que ce modèle de référence soit clairement explicite ou, comme c'est souvent le cas, qu'il demeure caché sous les mécanismes institutionnels, il n'en est pas moins omniprésent* (Artaud, 1989, 68-69).

Autant il serait insensé de nier l'existence d'un idéalisme dans l'acte d'évaluer des apprentissages, autant il serait inapproprié de s'en offusquer, d'en avoir honte ou de s'en culpabiliser. Cette attitude humaine d'idéaliser est une énergie qui nous a permis des réalisations intéressantes sur le plan des théories et des techniques en évaluation des apprentissages. Cette attitude influencera probablement encore les futures recherches. Quoi qu'il en soit, notre tâche n'est nullement de juger cette recherche d'un acte idéal d'évaluation des apprentissages. Il s'agit plutôt de regarder comment cet idéalisme a influencé nos conduites afin de mieux comprendre notre position actuelle et notre orientation future dans le domaine de l'évaluation des apprentissages.

2.2 L'émergence d'une série de mouvements

Ce qui est remarquable, c'est l'influence de vie qu'a procuré cet idéalisme dans l'acte d'évaluer des apprentissages. En effet, cette recherche d'un acte «idéal» d'évaluation des apprentissages a généré une série de mouvements qui indique l'évolution que prend l'évaluation en éducation. D'abord, il y a eu, dans nos recherches et dans nos pratiques, ce passage marquant d'une évaluation normative et certificative à une évaluation davantage régulatrice des apprentissages. L'évaluation est devenue un acte d'aide à la réalisation et à la maîtrise des apprentissages

prédéterminés et visés. Il y a donc eu une intensification pour l'évaluation dite formative qu'il ne faut pas confondre avec une évaluation formatrice.

Il serait intéressant de s'interroger sur ce qui a amené ce mouvement et inspiré les réflexions, les théories et les techniques des chercheurs, car c'est un mouvement qui a profondément marqué l'acte d'évaluer des apprentissages. Quelles motivations profondes ont influencé la conduite de ces chercheurs? À quels besoins ou désirs voulaient-ils répondre? Se pourrait-il que l'arrivée de ce mouvement et surtout l'intérêt manifesté par plusieurs chercheurs pour une évaluation plus formative ou plus soucieuse d'aider l'élève dans la maîtrise des apprentissages visés proviennent, entre autres, de cet ensemble de facteurs que représentent a) l'histoire personnelle de ces individus, c'est-à-dire leurs expériences vécues par rapport à l'évaluation comparative et normative qui a certes laissé des «empreintes» ou des «souvenirs refoulés» (Janov, 1992); b) une prise de conscience du non sens, pour la réalisation des apprentissages, d'actes évaluatifs uniquement à caractère comparatif et sélectif; c) la nécessité de donner un sens plus pédagogique à l'évaluation des apprentissages?

Ce dernier facteur pourrait expliquer l'apparition du mouvement qui vise à intégrer l'acte évaluatif à l'acte d'enseigner. Autrement dit, qui cherche à éliminer le fossé entre l'évaluation et l'enseignement et à faire en sorte que ces deux actes puissent mutuellement se compléter.

Il existe d'autres mouvements importants à signaler par rapport à l'évaluation des apprentissages. Il existe un mouvement qui s'intéresse à la qualité des instruments. Par exemple, des efforts considérables ont été déployés pour construire des instruments qui puissent recueillir, chez l'élève, des informations pertinentes, précises et fiables quant aux apprentissages visés. Il y a également eu des efforts pour construire des bulletins qui décriraient, le mieux possible, l'état des apprentissages de l'élève.

Depuis quelques années, il y a un mouvement qui cherche à impliquer davantage l'apprenant évalué dans la réalisation de ses apprentissages. Il s'agit de l'autoévaluation qui serait «au cœur même des processus de régulation d'un apprentissage» (Paquay, Allal et Laveault, 1990).

Un mouvement qui se poursuit lentement mais continuellement est celui de la «psychologie de l'évalué», c'est-à-dire qu'on se préoccupe du monde intérieur de l'évalué comme les effets de la «peur de l'évaluation» (Desruisseaux, 1977), les tensions générées par l'évaluation (Larson et Thienhaus, 1988), ou la recherche de moyens pour aider l'évalué à utiliser ses ressources et son pouvoir pour réussir à manifester l'état de ses apprentissages (Auclair, 1990; Desruisseaux et Devaux, 1984).

Enfin, il est impossible d'ignorer le mouvement, qui depuis plusieurs années, cherche à amener l'enseignant dans un «changement de son acte d'évaluation des apprentissages». Un changement dans le sens de parvenir à ce que l'acte hypothétiquement idéal puisse transparaître à travers la pratique évaluative de l'éducateur.

Trois constatations importantes se dégagent de cette rétrospective des mouvements engendrés en évaluation des apprentissages, constatations qu'il est possible de relever dans le schéma présenté à la page suivante. D'abord, ces mouvements rattachés à la recherche d'un acte idéal d'évaluation des apprentissages recherchaient trois formes d'impact : aider l'élève dans la maîtrise des apprentissages visés, construire et faire utiliser des instruments de qualité pour l'évaluation de ces apprentissages, et modifier sensiblement l'acte d'évaluer de l'enseignant.

Quant à la seconde constatation, il est intéressant de remarquer comment, à travers ces mouvements et cette recherche de l'idéalisme, l'éducateur a très peu été concerné et impliqué. Comme si l'évolution dans le domaine de l'évaluation des apprentissages s'était déroulée à l'extérieur de l'auteur principal des actes évaluatifs effectués envers l'élève. Il faut se demander si la pratique évaluative réelle de l'éducateur nous intéressait vraiment. Est-ce que la recherche d'un acte idéal d'évaluation des apprentissages nous aurait, inconsciemment, déconnectés de la pratique évaluative de l'éducateur? Pourtant, c'est cette pratique et non l'acte idéal qui a un impact effectif sur l'évalué. Il faut également ajouter qu'il est plus simple de travailler à l'élaboration d'un acte idéal d'évaluation que sur la pratique évaluative de l'éducateur.

Enfin, la dernière constatation à relever, c'est la nécessité de poursuivre ce qui a été accompli jusqu'à présent dans le domaine de l'évaluation des apprentissages, en s'engageant dans un autre mouvement. Il s'agit d'un mouvement qui pénètre davantage à l'intérieur de l'acte d'évaluer, un mouvement qui s'intéresse à la pratique réelle de l'éducateur et qui, par conséquent se préoccupe de faire une place primordiale à la personne de l'éducateur.

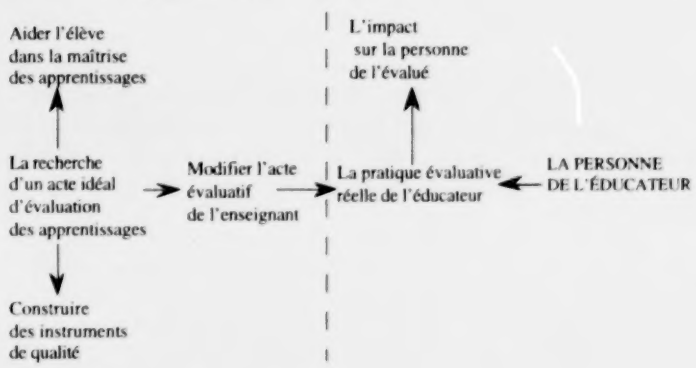


Schéma 1 – L'acte d'évaluer des apprentissages

3. FAIRE UNE PLACE À LA PERSONNE DE L'ÉDUCATEUR

Évidemment, on est en droit de se demander comment on va pénétrer dans ce monde plutôt méconnu de l'acte d'évaluer des apprentissages, là où réside la personne de l'éducateur. Dans le but d'ouvrir et d'approfondir quelque peu la réflexion, je propose de regarder et d'explorer les trois questions suivantes. Que signifie «faire une place à la personne de l'éducateur» dans l'acte d'évaluer des apprentissages? Qu'est-ce qui nous motive à faire une place à la personne de l'éducateur? De quelle manière faire une place à la personne de l'éducateur pour le changement de sa pratique évaluative?

3.1 Que signifie «faire une place à la personne de l'éducateur»?

Aussi simple que cela puisse paraître, il s'agit de prendre en considération la personne de l'éducateur et de l'inclure dans nos réflexions

et dans nos recherches. Plus concrètement, cela signifie, entre autres, s'intéresser à la nature de la personne de l'éducateur. Qui est cette personne qui vit en l'éducateur, auteur de nombreux actes conscients et inconscients d'évaluation? Quelle est l'histoire de cette personne en évaluation? C'est-à-dire quelles sont ses racines et ses expériences à propos de l'évaluation? Quelle est la nature de ses pensées, de ses sensations et de ses sentiments envers le phénomène de l'évaluation des apprentissages? Quelle perception de l'évaluation des apprentissages possède-t-elle? Qu'est-ce qui aurait pu influencer le développement de sa perception et comment cette dernière influence-t-elle sa pratique évaluative?

De quelle manière la personne de l'éducateur vit-elle le rapport qu'elle entretient consciemment et inconsciemment avec les demandes externes qui lui sont adressées ou qu'elle perçoit au sujet de son acte d'évaluer des apprentissages? Quel niveau de conscience a-t-elle de sa structure intérieure, de sa pratique évaluative et de leur relation? Quelle conscience a-t-elle de la nature et de l'impact de ses gestes formels et informels d'évaluation? Quel regard a-t-elle du rapport entre son monde intérieur et ses actes d'évaluation ainsi qu'entre ses transformations intérieures et le changement de sa pratique évaluative?

Lors de ses actes d'évaluation des apprentissages, comment la personne de l'éducateur se réfère-t-elle à certaines parties du monde intérieur, par exemple, à son centre interne d'évaluation, à son intuition, à sa subjectivité, à sa créativité, à ses habiletés intellectuelles? Comment se réfère-t-elle également à la théorie et technologie propre à l'évaluation des apprentissages?

Que fait-elle avec ses difficultés, ses malaises conflits, ses tensions... générés lors de certains actes d'évaluation? Quelle intégration a-t-elle faite de ses expériences d'évaluation? Et, enfin, où en est-elle dans l'appropriation de sa pratique évaluative ainsi que de sa propre formation en regard de l'évaluation des apprentissages?

Faire une place à la personne de l'éducateur en évaluation des apprentissages, c'est lui donner un espace afin qu'elle prenne la parole, qu'elle s'exprime par rapport à son vécu, autant au niveau de son monde intérieur qu'à celui de sa pratique évaluative, un espace qui lui permette de percevoir et de ressentir qu'elle est aussi importante dans

l'acte d'évaluer des apprentissages et dans le changement de sa pratique évaluative, et, sinon plus importante que la théorisation et la technologie qui ont surtout été au centre de nos discours et de nos études jusqu'à présent.

Donner un espace à la personne de l'éducateur afin qu'elle apprenne à descendre à l'intérieur d'elle-même et à effectuer les prises de conscience qui s'imposent pour rendre éducative sa pratique évaluative. L'acte d'évaluer est trop puissant et engendre trop d'expériences (empreintes) chez l'évalué pour l'accomplir avec inconscience, avec insensibilité et avec désinvolture. Voilà ce que signifie essentiellement «faire une place à la personne de l'éducateur» dans le domaine de l'évaluation des apprentissages.

Faire un premier pas dans la création d'une place à la personne de l'éducateur, c'est d'abord reconnaître et admettre que, jusqu'ici dans les réflexions et les recherches sur l'acte d'évaluer des apprentissages, on s'est davantage centré sur le raffinement d'un savoir théorique et technologique de même que sur la recherche d'un acte d'évaluation pensé, idéalisé et rêvé à l'extérieur de l'éducateur, d'une part, et également qu'on a cherché à ajuster son acte d'évaluation en fonction d'un acte idéal d'évaluation, d'autre part. Reconnaître et admettre que ce travail externe à l'éducateur nous a inévitablement conduit à le traiter davantage comme un «exécuteur» d'actes évaluatifs, comme un «moyen et non comme un sujet à qui on rend sa valeur et la dignité de l'être» (Abraham, 1984, 11).

Faire un second pas dans la création d'une place à la personne de l'éducateur en évaluation des apprentissages, c'est s'engager dans la recherche d'interventions qui favoriseraient, chez ce dernier, de prendre contact avec sa propre vie intérieure, là où réside la source de ses actes évaluatifs ainsi que son pouvoir de s'approprier et de changer sa pratique évaluative. C'est ce dont mes expériences personnelles et professionnelles m'ont amené à prendre conscience. «Les expériences nous ont convaincu de l'importance du subjectif. Le réveil, difficile, insécurisant du domaine de l'éducation ne peut se produire qu'à travers «son propre monde intérieur» (Abraham, 1972, 14).

Le schéma ci-dessous (schéma 2) tente d'offrir un aperçu, non exhaustif, de l'univers de la personne de l'éducateur dans le domaine de

l'évaluation des apprentissages. L'univers de la personne de l'éducateur est donc constitué de trois entités. En premier, cette personne possède une pratique évaluative observable, c'est-à-dire un ensemble de conduites,

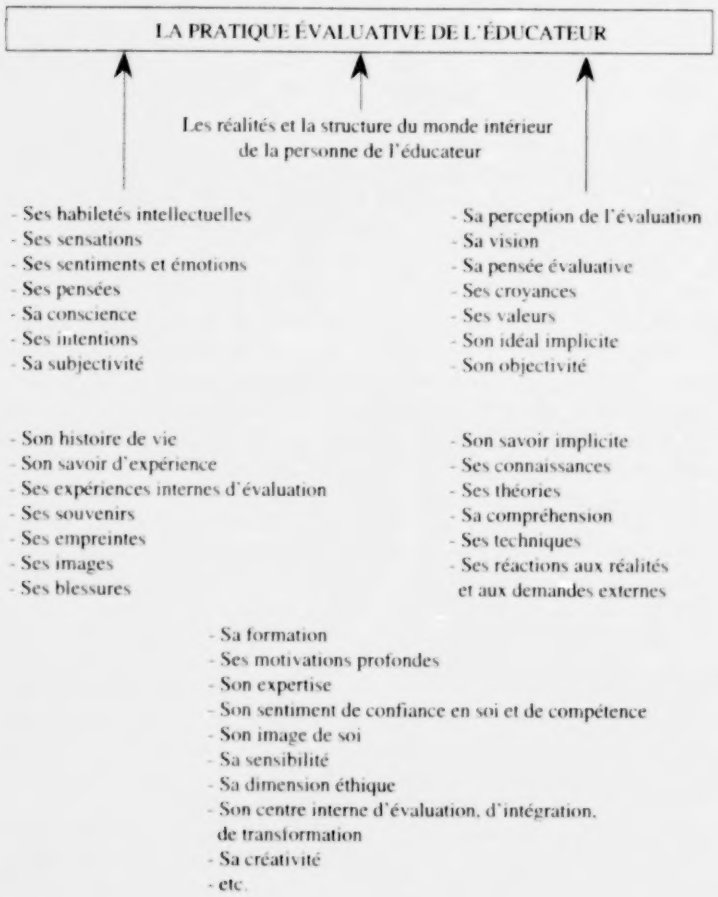


Schéma 2 – L'univers de la personne de l'éducateur dans l'évaluation des apprentissages

d'attitudes, d'actions spontanées et non spontanées, de gestes et d'actes évaluatifs à la fois conscients et inconscients, explicites et implicites, formels et informels. La personne de l'éducateur possède également une structure intérieure, c'est-à-dire un ensemble de réalités intériorisées qui se sont accumulées depuis l'enfance. Et enfin, comme troisième entité, il y a l'influence constante de sa structure intérieure sur sa pratique évaluative, une influence qui, d'habitude, se vit inconsciemment.

Il ne fait plus de doute que l'univers de la personne de l'éducateur dans l'évaluation des apprentissages est complexe, tout comme le phénomène de l'évaluation en soi, que la pratique évaluative de l'éducateur est subordonnée à sa structure intérieure et qu'il a la responsabilité de «prendre possession» de son univers et d'apprendre : a) à être en contact autant avec sa structure intérieure qu'avec sa pratique évaluative; b) à s'interroger sur la finalité, sur la puissance et sur l'impact de sa pratique évaluative sur le développement de l'Autre; c) à prendre conscience de la nature et de la dynamique de ses trois entités; d) à être à l'écoute de ses transformations intérieures et de leurs effets sur sa pratique évaluative; e) à apporter, à sa pratique évaluative, un changement qui émerge de l'intérieur.

Tout ce travail personnel s'installe dans le même sens de la recherche d'une autonomie personnelle et professionnelle de plus en plus recommandée : «il nous faut être autonomes vis-à-vis d'institutions en déclin, il nous faut être autonomes pour promouvoir notre désir» (Vassileff, 1992). Il s'agit d'un travail de «création de sa personne» :

> *L'être humain n'est pas donné à soi-même une fois pour toutes; au contraire, un travail essentiel illimité reste à faire : celui de se construire soi-même et de tenter de s'accorder à l'ampleur de la vie. Cette tâche, c'est celle de travailler à devenir soi-même, à chercher à composer avec les forces profondément divergentes qui nous habitent. Cette action de soi sur soi est un travail ardu, long et complexe et qui ne fait que commencer au seuil de l'âge adulte. C'est une éducation profonde, c'est-à-dire l'éducation véritable et sans fin pour chacun* (Gingras, 1986).

Il s'agit aussi d'un travail professionnel sur le plan de l'éthique, c'est-à-dire de l'interrogation d'un sujet sur la finalité de ses actes (Meirieu, 1991)

> *Interrogation qui le place d'emblée devant la question de l'Autre, car l'existence de l'Autre, chaque fois que j'agis et au sens propre des mots, «fait question» : est-ce que je le reconnais comme tel, dans sa radicale altérité, ou est-ce que j'en fais l'objet de mes manipulations pour servir à ma satisfaction? Dans tout ce que je dis, à travers toutes les décisions que je prends, au sein des institutions que je fréquente, est-ce que je permets à l'Autre d'être, face à moi, voire contre moi, un Sujet? Est-ce que j'accepte, en dépit des difficultés que cela comporte, de l'incertitude devant laquelle cela me place, des inquiétudes qui vont immanquablement surgir à chaque pas, de prendre ce risque? Telle est, pour nous, la question éthique fondamentale* (Meirieu, 1991, 11-12).

3.2 Qu'est-ce qui nous motive à «faire une place primordiale à la personne de l'éducateur»?

L'exploration de la question précédente a probablement mis en lumière certains motifs qui nous incitent à la nécessité de faire une place à la personne de l'éducateur dans le domaine de l'évaluation des apprentissages et du changement de sa pratique évaluative. J'aimerais examiner, plus spécifiquement, quatre ordres de motivations.

Le premier ordre de motivations concerne la nécessité, pour l'éducateur (auteur incontesté de ses actes évaluatifs) de travailler personnellement au développement de sa conscience par rapport à l'impact de sa pratique évaluative sur la croissance de l'élève, d'une part, et à la dynamique et à l'influence inconsciente de ses réalités intérieures sur sa pratique évaluative, d'autre part. Travailler à être davantage conscient de ce qu'on est comme personne, de la nature de sa pratique évaluative ainsi que de ce qui se cache dans son monde intérieur est au centre des responsabilités de tout éducateur.

Par expérience, je peux affirmer que le développement de sa conscience, en l'occurrence par rapport à l'évaluation des apprentissages, est facilité et stimulé lorsqu'on fait une place à la personne de l'éducateur et lorsque cette dernière perçoit et ressent qu'elle a une place.

Il est maintenant impossible de nier cette nécessité d'élargir sa conscience par rapport à la force que possède sa structure intérieure sur sa pratique évaluative. «Derrière l'action, avant même que les théories ne viennent justifier, expliquer, prédire, il y a la structure profonde de l'enseignant» (Paré, 1988, 5). La psychologie développementale de l'être humain nous permet de constater comment la structure profonde de tout individu joue constamment sur sa façon d'agir et d'intervenir (Bradshaw, 1992; Horney, 1955; Janov, 1992; Lowen, 1985, 1987, 1988, 1990; Miller, 1984, 1986; Progoff, 1984; Stettbacher, 1991). C'est également ce qu'Artaud (1985, 1989) a mis en évidence dans ses études sur l'intervention éducative, ainsi qu'Abraham (1972, 1984) dans ses recherches sur le monde intérieur de l'enseignant. Il y aurait certes des recherches à poursuivre dans le domaine de l'évaluation des apprentissages afin de saisir le rapport entre la structure interne de l'éducateur et sa pratique évaluative, un rapport très peu étudié et exploré jusqu'à présent.

Enfin, en regard avec ce premier ordre de motivations concernant le développement de sa conscience, il faut mentionner qu'il y a eu, au cours des dernières années, des chercheurs (Carbonneau et Trahan, 1980; Boisvert, 1981; Boisvert et Robillard, 1982; Desruisseaux, 1977; Trahan *et al.*, 1983) qui nous ont invité à prendre conscience de l'impact que pouvait avoir l'évaluation sur l'évalué. Il faut toutefois noter que, dans ces recherches, on parlait davantage de l'impact du phénomène de l'évaluation en général et non pas nécessairement de l'impact de la pratique évaluative de l'éducateur. De plus, on nous invitait à prendre conscience de l'évalué et non pas de l'univers de la personne de l'éducateur, tel que cela est décrit précédemment.

Le second ordre de motivations qui m'apparaît important de regarder concerne la notion du changement et, plus particulièrement, le «lieu de changement» de la pratique évaluative de l'éducateur. Comme il a été mentionné précédemment, depuis plusieurs années, il y a une

insistance remarquable afin que des enseignants modifient leur façon d'évaluer des apprentissages et qu'ils appliquent une démarche évaluative pensée et élaborée, entre autres, par des théoriciens et des techniciens de l'évaluation des apprentissages. La recherche d'un acte idéal d'évaluation, le développement théorique et technologique de l'évaluation formative ainsi que les séances de perfectionnement ont largement contribué à nourrir cette insistance.

Face à cette insistance de changement, plusieurs enseignants ont été confrontés à vivre un changement planifié, un changement dans lequel ils n'ont pas été impliqués, même s'il s'agissait de leur pratique évaluative. Il est possible de penser que certains enseignants ont dû vivre un changement plutôt dicté, voire forcé et imposé.

Ce changement planifié relatif au domaine de l'évaluation des apprentissages s'inscrit dans une conception mécaniste. Cette conception «conçoit le changement de l'homme comme programmé de l'extérieur par des personnes compétentes en fonction d'objectifs fixés à l'avance et socialement ou politiquement rentables» (Proulx, 1985, 23). À l'inverse, une «conception organique croit que c'est l'être en changement qui détermine et intègre lui-même son processus de changement. C'est donc un processus d'intégration personnelle chez l'individu» (Proulx, 1985, 23-24).

S'il avait été question d'utiliser une conception organique dans le domaine de l'évaluation des apprentissages, les enseignants auraient été davantage amenés à vivre un changement «émergent» par rapport à leur pratique évaluative, un changement qui vient d'abord de soi, de son intérieur, qui répond à un besoin et qui est pleinement ressenti, réfléchi et procure habituellement un plaisir, un changement qu'on a conscience d'éprouver et qu'on s'approprie pleinement. Bref, tout ce qui nous procure cette sensation d'en être l'auteur. Pour Proulx (1985), le changement «émergent se définit au fur et à mesure qu'on le vit».

> *On ne peut prévoir où notre recherche aboutira. Vivre un tel changement fait surgir la force intérieure qui existe dans chaque homme, il abandonne sa peur de lui-même, laisse tomber ses défenses intérieures et permet à sa personnalité profonde de*

> *surgir. De plus, ce type de changement n'est jamais fini puisqu'il se continue à travers une remise en question perpétuelle. Dès que je commence à m'accueillir, ma personnalité s'enrichit de tout ce qui arrive à ma conscience; je découvre que je suis un être en devenir. Je subis une mutation dans mes valeurs ce qui n'exclut pas des mouvements de retour en arrière. Mais l'état d'éveil et de conscience intérieure me rend attentive à ces difficultés et m'aide à y réagir* (Proulx, 1985, 21-22).

Il n'est certes pas question de contester la qualité du contenu ni les intentions visées dans le changement «programmé et planifié» qui a été vécu dans le domaine de l'évaluation des apprentissages. Toutefois, il s'avère important de s'interroger sur les conséquences d'un tel changement sur la personne de l'éducateur et en particulier sur le changement véritable et intégré que ce dernier a apporté à sa pratique évaluative.

On doit, entre autres, regarder l'effet qu'a pu avoir ce changement planifié, voire parachuté, d'évaluation des apprentissages sur le «système d'intervention» de l'éducateur, c'est-à-dire sur son ensemble organisé d'actions et de conduites, autant celles qui sont inconscientes que conscientes, ainsi que celles qui se situent entre les deux extrêmes de ce continuum. Chaque éducateur possède son propre système d'intervention intégré à l'intérieur de sa pratique éducative, un système issu, entre autres, de ses expériences, de ses apprentissages, de ses connaissances ainsi que de sa formation. Alors, l'insistance ou l'imposition d'adopter un changement planifié de l'extérieur quant à ses actions évaluatives ne peut pas être un geste neutre. C'est un geste qui a nécessairement des répercussions, si petites et inconscientes soient-elles, sur le reste du système d'intervention de l'éducateur et également sur sa structure intérieure : par exemple, affecter sa perception de l'évaluation, engendrer un certain état de tension, se sentir déconnecté de sa pratique évaluative, se désensibiliser de la personne de l'élève.

Jusqu'à présent, dans le domaine de l'évaluation des apprentissages, force nous est d'admettre que nous avons une connaissance très limitée concernant l'impact du changement planifié et parachuté sur le système d'intervention de l'éducateur. Comment pourrait-il en être au-

trement de cette méconnaissance, puisque la personne de l'éducateur n'a pas été au centre de nos préoccupations de recherches sur l'acte d'évaluer des apprentissages. À cet égard, Haramein (1979) avait déjà soulevé la nécessité de ne pas se limiter à un «entraînement théorique sur l'évaluation formative», mais également de s'intéresser au système d'intervention de l'enseignant pour favoriser, chez ce dernier, un changement intégré et général de sa pratique évaluative.

> *Il existe un certain hiatus entre l'image qu'a l'enseignant (ou n'importe quel être humain d'ailleurs) de son action et son action effective. On pourrait presque dire que le système d'intervention d'un enseignant est composé et de schèmes d'action à reproduction quasi automatique et d'interventions plus conscientes dont il a une représentation assez précise et, bien entendu, d'actions se situant entre ces deux extrêmes, à des distances plus ou moins grandes par rapport à la conscience. Tant et aussi longtemps qu'on ne disposera pas de connaissances plus précises sur ces différentes couches d'interventions, sur leurs structures, leurs mécanismes de régulation et leurs interactions, il serait vain de vouloir modifier efficacement le comportement des enseignants à partir, par exemple, d'un entraînement théorique portant sur l'évaluation formative. Car nous ne connaissons pas les conséquences que les modifications introduites dans la couche consciente ou réflexive pourraient avoir au niveau des autres couches d'intervention et nous ignorons, par conséquent, dans quelles conditions et suivant quelle stratégie nous parviendrons à réaliser un changement intégré et général en agissant uniquement sur les représentations rationnelles des enseignants* (Haramein, 1979, 165).

Lorsqu'on aborde les effets possibles d'un changement planifié, parachuté et imposé de l'extérieur à l'éducateur ou lorsque la personne de l'éducateur n'a pas été impliquée dans le changement de sa pratique, il faut regarder la possibilité de produire, chez ce dernier, une forme d'aliénation et de désappropriation de sa pratique : ce regard qui requiert du courage peut s'avérer dérangeant, mais est essentiel lorsqu'on se préoccupe de la personne de l'éducateur. Autrement dit, quand les choses viennent trop ou exclusivement de l'extérieur à soi-même, qu'elles sont

parachutées et surtout imposées, et lorsqu'on n'a pas été en mesure de se référer à sa structure interne (à sa connaissance implicite, ses expériences, son expertise, son centre interne d'évaluation...) pour situer ces choses externes et les intégrer dans son univers, il y a alors danger de vivre une forme d'aliénation et de désappropriation : une aliénation dans le sens que l'éducateur «n'existe plus par lui-même mais par le regard d'autrui, par ce qu'on lui dicte et impose» (Artaud, 1985, 37). Pour Paré, il s'agit d'une «coupure profonde avec ses racines».

> *C'est-à-dire est-ce que ma connaissance est issue de mon expérience et de mes racines ou si elle prend prise dans l'idée de monsieur ou madame Untel, qui est venu me dire ce que je devrais faire et à quelle condition je serais correct? Si on vient me dire cela, que je trouve ça intéressant et que je rentre là-dedans, je risque de me mettre à devenir ça et de me distancer de plus en plus de moi, de mon expérience intérieure. Tandis que si je descends dans ma connaissance à moi, dans ma difficulté à moi, il est possible que ces théories viennent à point, juste pour me donner un langage pour parler. Alors il y a une connaissance implicite, il y a une connaissance explicite et il me paraît qu'il faut que chaque enseignant descende tranquillement vers sa connaissance implicite* (Paré, 1989).

Le fait d'exécuter, inconsciemment et consciemment, des actes d'évaluation des apprentissages en fonction des demandes externes et non en référence à soi-même, le fait également «d'évacuer les acteurs impliqués dans la démarche de formation (en évaluation des apprentissages) et de centrer celle-ci sur la transmission de théories d'une part et la construction d'une pratique hypothétique à partir d'elles, d'autre part» (Haramein, 1989), cela risque de provoquer, chez l'éducateur, une sensation d'être désapproprié et dépersonnalisé de sa pratique évaluative. Comment est-il possible de ne pas éprouver une telle sensation de désappropriation et de dépersonnalisation lorsque la personne de l'éducateur est exclue, et s'en exclut également, du changement de sa pratique évaluative?

Être désapproprié et dépersonnalisé de sa pratique évaluative, c'est ressentir que ses actions et ses interventions évaluatives ne pro-

viennent pas de soi, qu'elles sont la possession du «regard d'autrui», qu'elles sont dirigées par l'extérieur à soi; c'est avoir cette sensation que «quelque chose nous échappe» et qu'on devient coupé de sa pratique évaluative. Ce qui est paradoxal, c'est que, malgré cette sensation, l'éducateur demeure l'auteur responsable de chacun de ses actes évaluatifs. C'est d'ailleurs ce qui engendre un état de tension difficile à vivre et que graduellement chacun de ses actes d'évaluation devient difficile, voire pénible à effectuer.

Il faut alors cesser d'induire le changement de l'extérieur, comme le préconisent Paré et Auclair (1988) et permettre à l'éducateur de comprendre qui il est, comment et pourquoi il effectue sa pratique évaluative. «Nous croyons que toute personne est capable d'entrer en contact avec les motivations profondes qui la font intervenir de telle ou telle façon, et que, sans une conscience de ces raisons, de ces motivations essentielles, elle ne peut transformer significativement son intervention» (Paré et Auclair, 1988, 31).

C'est dans cette perspective qu'un groupe de chercheurs du département de psychopédagogie de l'Université de Montréal (Fournier, Gingras, Haramein et Hétu) ont fondé le labo-TIPÉ (Laboratoire de recherche sur la transformation intérieure et la pratique éducative). Ces chercheurs mettent l'emphase sur la nécessité, pour l'éducateur, de prendre conscience autant de son monde intérieur que de sa pratique éducative. Ils travaillent donc à clarifier diverses facettes des rapports entre la transformation intérieure et la pratique éducative en aidant l'éducateur à prendre en charge sa propre formation et le changement de sa pratique éducative. «Le changement et le renouvellement de la pratique éducative passent en premier lieu, sinon de façon exclusive, par le ressourcement et la transformation intérieure de la personne des éducateurs» (Gingras, 1992).

Cette nouvelle avenue de recherche apparaît excessivement intéressante, appropriée, voire essentielle, dans le domaine de l'évaluation des apprentissages. En effet, elle va dans le même sens du mouvement qu'il nous faut amorcer dans le domaine de l'évaluation des apprentissages, soit celui de faire une place primordiale à la personne de l'éducateur pour le changement de sa pratique évaluative.

Cette place primordiale à faire à la personne de l'éducateur est dictée, non seulement par la nécessité, pour l'éducateur, de travailler au développement de sa conscience et de s'approprier le changement de sa pratique évaluative, comme il a été discuté précédemment. Il existe également un autre ordre de motivations qui doit être pris en considération dans cette nécessité de faire une place à la personne de l'éducateur : il s'agit de «l'évolution des recherches» dans le domaine de l'évaluation des apprentissages.

Après s'être doté de théories, de techniques et d'instruments convenables par rapport à la recherche d'un acte idéal d'évaluation – certes qu'un travail demeure encore à faire dans chacun de ces domaines –, le temps est venu d'amorcer un autre mouvement, soit de passer du savoir théorique au savoir d'expérience, de la technologie à la pratique évaluative réelle, de la mécanique à la personne de l'éducateur, c'est-à-dire d'aller voir et explorer le côté où prend forme et se vit l'acte réel d'évaluer des apprentissages. Passer du côté qui nous est inconnu soit la personne de l'éducateur et nous ouvrir à des approches et à des méthodes de recherche qui s'intéressent aux réalités de la personne de l'éducateur dans l'acte d'évaluer des apprentissages.

Il est possible de voir, à travers certaines recherches dans le domaine de l'évaluation des apprentissages, se dessiner ce mouvement en approchant doucement certaines dimensions de la personne de l'éducateur. Parmi ces dimensions, il y a le concept de l'équité chez l'évaluateur (Bélair, 1990) de même que son jugement moral (Bélair, 1988), la mesure des croyances et attitudes des enseignants (Burelle *et al.*, 1991; Fontaine, 1988; Fontaine et Trahan, 1990a, 1990b; Gadbois *et al.*, 1991), la subjectivité de l'évaluateur (Weiss, 1986), le comportement d'évaluation de l'enseignant (Noizet et Caverni, 1978). Ces premiers pas nous invitent à pénétrer davantage dans l'univers de la personne de l'éducateur, entre autres, pour le changement de sa pratique évaluative.

Je terminerai l'exploration de cette question sur le pourquoi de «faire une place à la personne de l'éducateur» par un propos plus personnel qui représente le quatrième ordre de motivations). Il ne m'est plus possible d'imaginer une quelconque éducation sans faire une place à la

personne; ceci ne signifie pas, pour autant, que je ne suis pas préoccupé par l'apprentissage. Au contraire! Pour moi, il ne peut y avoir apprentissage véritable, significatif, pleinement consenti et intégré sans la participation totale de la personne de l'apprenant. C'est ce que je découvre en retraçant l'histoire de mes apprentissages. Mes plus beaux et bons apprentissages, pas nécessairement les plus faciles, ont été réalisés lorsque ma personne a été entièrement impliquée, lorsqu'elle a senti qu'on se préoccupait autant d'elle que de l'apprentissage visé, créant ainsi une sensation d'être respecté et aidé dans mes besoins, dans mes désirs et dans ma capacité d'apprendre.

Je réalise comment ces expériences sont des racines importantes sur lesquelles s'appuient cette croyance profonde en la nécessité de faire une place primordiale à la personne de l'éducateur, entre autres, dans le domaine de l'évaluation des apprentissages. D'ailleurs, ces deux personnes, celle de l'apprenant et celle de l'éducateur, ne se dissocient pas. Comme éducateur, j'ai également besoin que ma personne ait une place primordiale afin que ma pratique évaluative soit éducative.

Il m'est impossible de faire taire la voix de cette personne qui m'habite depuis l'enfance, une voix qui revendique qu'on fasse une place primordiale à ma personne d'éducateur et non seulement à l'acte idéal d'évaluer des apprentissages visés et à l'évalué. Il ne s'agit nullement de prendre leur place, mais uniquement avoir ma place comme éducateur, ce qui n'est pas toujours évident en éducation. Dans cette voix intérieure, j'entends également celle d'autres éducateurs lui m'ont manifesté et partagé comment ils leur étaient parfois difficile, voire pénible, de supporter quotidiennement le poids de l'évaluation, car il faut bien se dire que l'évaluation des apprentissages peut devenir une lourdeur, un lieu de tension et de conflits dont certains auteurs (Hannoun, 1975; Perrenoud, 1984) osent parler. Des éducateurs m'ont également exprimé leur colère et leur déception de constater et surtout de ressentir à quel point leur personne avait peu de place en évaluation des apprentissages et dans le changement de leur acte d'évaluer.

Aujourd'hui, je sais, par expérience, comment il est plus agréable de vivre sa pratique évaluative d'une manière éducative et d'y apporter des changements intégrés lorsque la personne de l'éducateur a une place

primordiale. J'en arrive à la conclusion qu'il faut, non seulement faire une place à la personne de l'éducateur en évaluation des apprentissages, mais surtout chercher à faire soi-même une place à sa propre personne.

3.3 Comment faire une place à la personne de l'éducateur pour le changement de sa pratique évaluative?

Il m'est plus difficile à aborder cette question que les deux précédentes. Cette difficulté provient surtout du sentiment de méfiance et de prudence qui s'installent par rapport au comment. Il m'apparaît inévitable que parler du comment c'est risquer de s'enliser dans l'aspect méthodique, technique, mécanique, dans le truc ou la recette... ce que je refuse et qui serait inapproprié dans cette recherche humaine sur la place à faire à la personne de l'éducateur dans le domaine de l'évaluation des apprentissages.

Pour moi, le comment doit être utilisé comme une interrogation personnelle récupérée par chaque individu, une interrogation qui l'aide à découvrir ou à créer sa propre manière de faire selon ce qui fait du sens pour lui-même, le point de référence pour tout changement. D'ailleurs, il aurait été impensable de s'engager dans le comment, qui représente l'aspect davantage pratique, sans au préalable s'attarder à la «signification» et au «pourquoi» faire une place à la personne de l'éducateur, sans au préalable saisir cette signification et ce pourquoi. À quoi peut servir l'application d'une méthode ou d'une technique quelconque lorsqu'on n'a pas pris le temps de préciser ce qui en motive et en justifie son utilisation.

Tout ce qui a été mentionné et examiné jusqu'à présent fait ressortir que, pour favoriser un changement intégré de sa pratique évaluative, les deux processus suivants doivent être privilégiés auprès de la personne de l'éducateur : une prise de contact avec l'univers de sa personne et une prise de parole.

Avant de s'introduire à l'intérieur de ces deux processus, quelques précisions s'imposent. D'abord, les approches, les techniques et les interventions qui seront présentées s'inscrivent dans un changement émergent, c'est-à-dire dans un changement intégré qui provient de la per-

sonne de l'éducateur. Il est important aussi de se rappeler que «faire une place à la personne de l'éducateur», c'est permettre à cette personne d'avoir un espace pour favoriser une prise de contact avec sa pratique évaluative, avec ses réalités internes et leurs connexions. Malgré l'aide que procure un environnement externe qui se centre à faire une place à la personne de l'éducateur, il n'en demeure pas moins que la création de cet espace est la responsabilité première de l'éducateur. Apprendre à faire une place de choix à sa propre personne pour le changement de sa pratique évaluative est un travail personnel qui ne s'impose pas de l'extérieur. Enfin, cette place primordiale à faire à la personne de l'éducateur n'est réalisable qu'à la condition de croire au désir profond, de chaque éducateur, de vouloir se connaître, se comprendre et se transformer, de croire également à la dimension significative de sa pratique évaluative et de ses réalités intérieures ainsi qu'à la richesse de l'expérience humaine.

3.4 Le processus de «prise de contact» avec l'univers de sa personne

Cet acte est le point de départ pour amorcer un changement intégré de sa pratique évaluative. Il ne peut y avoir un changement qui vient de l'intérieur si l'éducateur n'entre pas en relation avec l'univers de sa propre personne, s'il ne dirige pas un regard dénudé de jugement de valeur sur sa pratique évaluative et ses réalités intérieures.

Autrement dit, ce processus de prise de contact conduit dans un travail d'intimité avec soi-même, d'intériorisation et d'introspection. Si le recours à l'intériorisation et à l'introspection est valorisé et profitable pour le développement et le pour fonctionnement des «gestes mentaux» chez l'élève (de la Garanderie, 1984, 1989, 1991; Taurisson, 1988) ou pour l'actualisation de son potentiel créateur (Fugitt, 1984; Galyean, 1986), alors il devrait en être également ainsi pour l'éducateur engagé dans une recherche éducative de sa pratique évaluative. Prendre contact avec l'univers de sa personne s'inscrit dans une «recherche d'introspection, de méditation et d'analyse réflexive sur la nature et la signification de son expérience» (Moustakas, 1990, 11).

Dans ce processus de prise de contact, la personne de l'éducateur établit un rapport subjectif avec son univers, rapport qui devient son

instrument pour l'exploration de sa pratique évaluative et de ses réalités internes, et elle pose un regard objectif sur ce qu'elle est en train de vivre par rapport à l'évaluation des apprentissages. L'éducateur est alors engagé dans une forme de recherche-formation (Haramein, 1991) et de recherche heuristique (Moustakas, 1990). Ce rapport subjectif et ce regard objectif se réalisent à l'aide de différentes approches centrées sur la personne : le *focusing* (Gendlin, 1984, 1986), la rétrospective (Levin, 1992), l'histoire ou le récit de vie (Ditisheim, 1991), le journal intime intensif (Progoff, 1984), le retour aux racines de son enfance (Bradshaw, 1992), l'exploration d'une expérience pénible à vivre (Stettbacher, 1991), l'utilisation des principes et des techniques de la psychosynthèse (Paré, 1984). Ce sont des approches qui font une place à la personne de l'éducateur en privilégiant le processus de prise de contact avec son univers.

3.5 Processus de «prise de parole»

Le travail d'intériorisation et d'introspection doit être accompagné d'une prise de parole par la personne de l'éducateur. Il ne s'agit pas de prendre la parole de l'acte idéal d'évaluation, comme cela se produit malheureusement trop souvent dans le domaine de l'évaluation des apprentissages : une parole théorique et technique sans rapport nécessaire avec la personne de l'éducateur. Le processus de prise de parole à privilégier ne consiste pas à parler du phénomène de l'évaluation sans rapport avec sa pratique évaluative. Il ne consiste également pas à décrire, d'une manière superficielle et déconnectée, quelques-unes de ses interventions évaluatives dans lesquelles, plus souvent qu'autrement, la personne de l'éducateur n'est pas impliquée; ceci reflète un discours non ressenti et non intégré.

Le processus de prise de parole dont il est question pour le changement de sa pratique évaluative va dans le sens de prendre sa parole. Il s'agit d'une parole personnelle, unique et intégrée au processus de prise de contact, d'une parole qui vient de soi et qui parle de soi ainsi que des rapports et des connexions entre sa pratique évaluative et ses réalités intérieures.

Cette prise de parole a comme objectif de se libérer afin de vivre pleinement un changement intégré de sa pratique évaluative. Se libérer dans le sens de toucher et de défaire ses nœuds de préoccupations, c'est-à-dire les conflits, les tensions que l'éducateur peut éprouver et retenir inconsciemment par rapport à l'évaluation des apprentissages, car se sont des nœuds qui peuvent entraver chez lui l'émergence d'un changement de sa pratique évaluative. Se libérer signifie aussi s'exprimer sur son univers, parler de ses expériences, de sa perception de l'évaluation ou de sa pratique réelle d'évaluation. Plusieurs auteurs (Abraham, 1972; Artaud, 1985; Binder, 1975; Lowen, 1987; Rogers, 1969) insistent sur l'importance de s'exprimer, donc de prendre la parole, pour le développement de sa personne. Abraham (1972) dénonce le «silence» dans lequel s'emprisonne l'éducateur, un silence qui a un effet sur l'univers de sa personne. «C'est une population qui ne se manifeste pas. [...], ils n'expriment pas à haute voix leurs revendications, leurs besoins de formation, ce qu'ils pensent de leur situation ambiguë» (p. 13).

En plus de travailler sur soi, la prise de parole ouvre la voie à l'interaction et à la communication avec l'élève; c'est ce que certains auteurs (Bélair, 1991; Bourassa et Leblanc, 1991) préconisent comme mode d'évaluation des apprentissages. Une interaction et une communication qui permettent également à l'élève de prendre la parole, c'est-à-dire de parler de lui, de s'exprimer sur son monde intérieur; ce qui favorise l'éducation à l'évaluation. «C'est parce qu'on prend la parole, sa parole, qu'on a quelque chose de personnel, d'enraciné et de singulier à dire, que la communication et l'interaction deviennent possibles et, inversement, c'est grâce à celles-ci que la singularité émerge et se développe» (Haramein, 1991). C'est donc deux personnes qui prennent une place dans l'acte d'évaluer des apprentissages : la personne de l'élève et celle de l'éducateur. De plus, le changement émergent de sa pratique évaluative n'est pas seulement favorisé par la prise de parole de l'éducateur; la prise de parole de l'élève contribue, sans aucun doute, à ce changement.

Bien sûr que d'autres réflexions et études s'imposent au niveau du comment faire une place à la personne de l'éducateur pour le changement de sa pratique évaluative. Mais il ne fait aucun doute que la prise de contact avec son univers et la prise de parole sont des processus qui contribuent à faire une place à la personne de l'éducateur.

CONCLUSION

La réflexion sur la place primordiale à la personne de l'éducateur dans le domaine de l'évaluation des apprentissages et particulièrement pour le changement de sa pratique évaluative ne fait que commencer, mais voilà qu'elle s'annonce déjà comme une nécessité. L'évaluation ne peut devenir un acte d'éducation et d'aide au développement de l'individu si la personne de l'éducateur ne prend pas sa place. L'éducation est une affaire de personne à personne.

La réflexion sur la place de la personne de l'éducateur dans le domaine de l'évaluation des apprentissages doit se poursuivre afin que l'acte évaluatif devienne complètement intégré à l'acte pédagogique. En effet, le domaine de la pédagogie et plus particulièrement de la psychopédagogie se centre de plus en plus à ce que la personne de l'éducateur prenne sa place dans le changement de sa pratique éducative. Il serait impensable que la personne de l'éducateur ait une place primordiale dans son acte éducatif et que ce soit l'inverse dans son acte évaluatif.

Il est évident que cette réflexion soulève des inquiétudes et des insécurités. On peut même se demander ce qui va se passer si chaque éducateur fait une place à sa personne. C'est également la même question que l'éducateur peut se poser par rapport aux élèves. Que va-t-il se passer si chaque élève fait une place primordiale à sa personne pour la réalisation de ses apprentissages?

La seule réponse qui jaillit, c'est ma conviction profonde à l'importance de faire une place primordiale à la personne de l'éducateur pour le changement de sa pratique évaluative. Il me semble que cela ne peut nous engager dans une seule voie : celle d'une pratique évaluative plus éducative.

RÉFÉRENCES

ABRAHAM, A. (1989).
Une approche centrée sur la personne de l'enseignant. *Éducation et pédagogie, 1*, 23-32.

ABRAHAM, A. (1984).
L'enseignant est une personne (introduction). Paris : ESF.

ABRAHAM, A. (1972).
Le monde intérieur des enseignants. Paris : Épi.

ABRAHAM, A. *et al.* (1984).
L'enseignant est une personne. Paris : ESF.

ALLAL, L., CARDINET, J. et PERRENOUD, P. (1979).
L'évaluation formative dans un enseignement différencié. Berne : Peter Lang.

ARTAUD, G. (1989).
L'intervention éducative. Ottawa : Presses de l'Université d'Ottawa.

ARTAUD, G. (1985).
L'adulte en quête de son identité. Ottawa : Presses de l'Université d'Ottawa.

ARTAUD, G. (1982).
La crise d'identité de l'adulte et ses incidences éducatives. *Revue des sciences de l'éducation, VIII*(3), 463-480.

AUCLAIR, M. (1990).
Une démarche de diminution des tensions. *Intégration, 6*, 25-29.

BÉDARD, R. (1984).
Les modèles de développement de l'adulte. *Revue des sciences de l'éducation, X*(3), 447-466.

BÉDARD, R. (1983).
Crise et transition chez l'adulte dans les recherches de Daniel Levinson et de Bernice Neugarten. *Revue des sciences de l'éducation, IX*(1), 107-126.

BÉDARD, R. (1981).
Recherches en psychologie de l'adulte. *Revue des sciences de l'éducation, VII*(3), 393-416.

BÉLAIR, L. (1991).
Faire participer l'élève à l'évaluation de ses apprentissages (Préface : *Les vertus du courage ou les risques de la profession?*). Chomedey : Beauchemin.

BÉLAIR, L. (1990).
Évaluation: comparer en toute équité. *Mesure et évaluation en éducation, 13*(2), 23-45.

BÉLAIR, L. (1988).
L'influence des histoires utilisées dans les entrevues sur l'estimation du stade de jugement moral des adolescents et des adolescentes. *Mesure et évaluation en éducation, 10*(4), 33-57.

BINDER, J. (1975).
The discovery of personnes : The assessment of feeling expressiveness. California : University of California.

BOISVERT, M.T. (1981).
L'évaluation sur le plan affectif : état de la question. *Mesure et évaluation en éducation, 4*(2), 46-51.

DESRUISSEAUX, J.C. et DEVAUX, C. (1984).
Effets des techniques de relaxation sur la performance en milieu scolaire. *Mesure et évaluation en éducation, 7*(3), 25-44.

DITISHEIM, M. (1991).
Le travail de l'histoire de vie : formation ou thérapie? *Revue des sciences de l'éducation, XVII*(2), 295-304.

DOYON, C. et JUNEAU, R. (1991).
Faire participer l'élève à l'évaluation de ses apprentissages. Chomedey : Beauchemin.

FONTAINE, F. (1988).
Mesure critériée des croyances des futurs maîtres à l'égard de l'évaluation des apprentissages. Montréal : Université de Montréal, Faculté des sciences de l'éducation.

FONTAINE, F. et TRAHAN, M. (1990a).
Pour une mesure des croyances dans un contexte de formation. Mesure et évaluation en éducation, 12(4), 5-22.

FONTAINE, F. et TRAHAN, M. (1990b).
Analyse du domaine des croyances reliées à l'évaluation des apprentissages. *Mesure et évaluation en éducation, 13*(1), 5-28.

FUGITT, E. (1984).
C'est lui qui a commencé le premier. Sainte-Foy : Centre d'intégration de la personne.

GADBOIS, L. *et al.* (1991).
Un instrument de mesure des croyances et attitude des enseignants à l'égard des pratiques d'évaluation formative des apprentissages de leurs élèves. *Mesure et évaluation en éducation, 14*(2), 5-23.

GALYEAN, B.C. (1986).
Visualisation, apprentissage et conscience. Sainte-Foy : Centre d'intégration de la personne.

GENDLIN, E.T. (1984).
Focusing : au centre de soi. Montréal : Le jour.

GINGRAS, J.M. (1992).
La transformation de la pratique éducative passe par celle des éducateurs. *Forum.*

GINGRAS, J.M. (1986).
Histoire de vie et expérience de création. Communication présentée au 54e congrès de à l'Association canadienne française sur l'avancement de la science.

HANNOUN, H. (1975).
Les conflits de l'éducation. Paris : ESF.

HARAMEIN, A. (1991).
La personne, trait d'union entre la recherche et la formation. *In* Bordeleau, L.G., Brabant, M, Cazabon, B, Desjardins, F. et Leblanc, R. (éd.), *Libérer la recherche en éducation – Actes du 3e congrès des sciences de l'éducation de langue française du Canada* (p. 367-373). Ottawa : Presses de l'Université d'Ottawa.

HARAMEIN, A. (1989).
Savoir académique et pratique professionnelle : une interaction sans acteur. *In* Roy, G.R. (éd.), *Contenus et impacts de la recherche universitaire en sciences de l'éducation – Actes du 2e Congrès des sciences de l'éducation de langue française du Canada* (p. 363-367). Sherbrooke : Université de Sherbrooke, Éditions du CRP.

HARAMEIN, A. (1979).
L'évaluation formative et le système d'intervention du maître. *In* Allal, J., Cardinet, J. et Perrenoud, P. (éd.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

HORNEY, K. (1955).
Nos conflits intérieurs. Paris : L'arche.

HUBERMAN, M. (1989).
La vie des enseignants. Évolution et bilan d'une profession. Neuchâtel : Delachaux et Niestlé.

HUBERMAN, M. (1984).
L'enseignant est une personne. Paris : ESF.

JANOV, A. (1992).
Le nouveau cri primal. Paris : Presses de la Renaissance.

LARSON, E. et THIENHAUS, O. (1988).
The development of self-confidence in psychiatric residents : A survey. *Journal of Psychiatric Education, 12*(1), 44-50.

LEVIN, R.F. (1992).
La rétrospective, un processus naturel. *Intégration*, (13), 10-25.

LOWEN, A. (1988).
La peur de vivre. Paris : Épi.

LOWEN, A. (1987).
Gagner à en mourir. Paris : Hommes et Groupes.

LOWEN, A. (1985).
Le corps bafoué. Montréal : France-Amérique.

LOWEN, A. (1980).
Le cœur passionnément. Paris : Tchou.

MEIRIEU, P. (1991).
Le choix d'éduquer. Paris : ESF.

MILLER, A. (1986).
L'enfant sous terreur. Paris : Aubier.

MILLER, A. (1984).
C'est pour ton bien. Paris : Aubier.

MOYNE, A., ARTAUD, J. et BARLOW, M. (1988).
Formation et transformation de l'enseignant. Vers une pratique renouvelée de la pédagogie. Lyon : Chronique sociale.

PARÉ, A. (1991).
Analyse évolutive d'une pratique pédagogique. *Intégration*, (10), 23-40.

PARÉ, A. (1989).
Changement de personnel et pratique pédagogique. Conférence donnée au VI[e] colloque du Goéland, Université de Montréal.

PARÉ, A. (1988).
Analyse évolutive d'une pratique pédagogique. *Recherches psychopédagogiques, 1*(2), 5-16.

PARÉ, A. (1984).
Psychosynthèse et éducation (introduction à l'édition québécoise du volume : C'est lui qui a commencé le premier). Sainte-Foy : Le Centre d'intégration de la personne.

PARÉ, A. (1983).
L'acte d'enseigner, geste d'art, source de perfectionnement. *Prospectives*, 175-1991.

PARÉ, A. et AUCLAIR, M. (1988).
Vidéoscopie, visualisation et supervision pédagogique. *Recherches psychopédagogiques, 1*(2), 31-51.

PAQUAY, L., ALLAL, L. et LAVEAULT, D. (1990).
L'auto-évaluation en question(s). Propos pour un débat. *Mesure et évaluation en éducation 13*(3), 5-26.

PERRENOUD, P. (1991).
Pour une approche de l'évaluation formative. *Mesure et évaluation en éducation, 13*(4), 49-81.

PERRENOUD, P. (1984).
La fabrication de l'excellence scolaire. Genève : Droz.

PERRENOUD, P. (1979).
Des différences culturelles aux inégalités scolaires : l'évaluation et la norme dans un enseignement indifférencié. *In* Allal, L., Cardinet, J., Perrenoud, P. (éd.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

POURTOIS, J.P. (1978).
L'évaluation de l'apprentissage ou l'apprentissage de l'évaluation. *In* Allal, L., Cardinet, J. et Perrenoud, P. (éd.), *L'évaluation formative dans un enseignement différencié*. Berne : Peter Lang.

PROGOFF, I. (1984).
Le journal intime intensif. Montréal : Éditions de l'homme.

PROULX, P. (1985).
Analyse du changement de relation éducative selon une approche heuristique. Mémoire présenté à la Faculté des sciences de l'éducation, section préscolaire-primaire, Université de Montréal.

MOUSTAKAS, C. (1990).
Heuristic research. Newbury (CA) : Sage.

NOIZET, G. et CAVERNI, J.P. (1978).
Psychologie de l'évaluation scolaire. Paris : PUF.

ROBILLARD, R. (1990).
Apprendre à utiliser son intériorité dans son acte d'évaluation. Communication présentée à la 13e sessions d'étude de l'ADMEE.

ROGERS, C. (1969).
Le développement de la personne. Paris : Dunod.

SCALLON, G. (1988a).
L'évaluation formative des apprentissages. Québec : Les Presses de l'Université Laval.

SCALLON, G. (1988b).
Plaidoyer pour une méthodologie instrumentée d'évaluation formative. *Mesure et évaluation en éducation, 11*(1), 43-55.

STETTBACHER, J.K. (1991).
Pourquoi la souffrance. Paris : Aubier.

TAURISSON, A. (1988).
Le gestes de la réussite en mathématiques à l'élémentaire. Montréal : Agence d'Arc.

TRAHAN, M. *et al.* (1983).
Interventions évaluatives : impact sur les évalués. *Mesure et évaluation en éducation, 6*(2), 69-78.

VASSILEFF, J. (1992).
Histoires de vie et pédagogie du projet. Lyon : Chronique sociale.

WEISS, J. (1986).
La subjectivité blanchie? *Mesure et évaluation en éducation, 9*(1), 45-68.

ÉTUDE DE LA VALIDITÉ DES ÉPREUVES DE FRANÇAIS ÉCRIT DE SIXIÈME ANNÉE DU PRIMAIRE DE JUIN 1990 ET DE JUIN 1991 EN CONFORMITÉ AVEC LE PROGRAMME DE FRANÇAIS LANGUE MATERNELLE DU MINISTÈRE DE L'ÉDUCATION DU QUÉBEC

RÉJEAN AUGER
Université du Québec à Montréal

INTRODUCTION

En 1986, le ministère de l'Éducation a rendu obligatoire une épreuve de français écrit pour tous les élèves de sixième année du primaire. Il n'existe pas de directive à l'effet que l'épreuve soit un critère de promotion des élèves mais il est fort probable que le résultat à l'épreuve représente un poids important dans la prise de décision. Le développement d'une grille d'appréciation pour évaluer la production écrite fut influencé par le courant de l'interprétation critérielle et par le besoin de mieux connaître les forces et les faiblesses des élèves à l'échelle du Québec (Ministère de l'Éducation, 1986). Cela a donné naissance à une échelle descriptive plus explicite des comportements attendus d'un élève dans un contexte d'évaluation sommative.

La préoccupation de répondre aux demandes du milieu scolaire et d'assurer une similitude entre la pratique des commissions scolaires en matière d'évaluation sommative de la production écrite et celle du ministère de l'Éducation a donné lieu, d'une année à l'autre, de la part du Ministère, à des façons différentes de noter et d'interpréter les résultats de la mesure. Ces modifications apportées, depuis 1990, sont plutôt mineures en ce qui concerne la forme et le fond, tellement qu'aucune étude, à ce jour, n'a examiné l'impact des changements. Le but recherché par cet écrit est d'examiner la validité globale de telles propositions et, plus particulièrement, ce qui concerne la classification des élèves.

À partir des propositions ministérielles à l'intérieur des épreuves de français écrit, une problématique est décrite. Un cadre d'analyse pour l'examen de la validité est proposé. Ce cadre utilise un concept élargi de validité intégré à une démarche évaluative où les résultats observés à la suite de diverses analyses sont appréciés en rapport avec le but visé, l'instrument de mesure utilisé, les interprétations proposées et les décisions de classement. C'est pourquoi la validité de contenu, de construit, discriminante, reliée aux critères décisionnels, l'équivalence des interprétations reliée à l'utilisation de deux échelles de mesure et la pondération de chacun des critères de correction font parties de l'examen de la validité.

1. PROBLÉMATIQUE

Dans le but d'harmoniser l'épreuve écrite à un contexte similaire aux pratiques scolaires, le Ministère propose un choix entre deux notations et une interprétation à partir des résultats de l'épreuve. Cette décision a été prise en vue de tenir compte des bulletins conventionnel et descriptif en vigueur au Québec. Cela s'est traduit, pour les épreuves de production écrite, dans une première notation où la note finale sur cent est constituée de la somme des notes pondérées à chacun des dix critères et dans une deuxième notation où la note finale correspond à la somme des critères réussis sur dix. La réussite à chacun des critères correspond à une dichotomie (réussite ou échec) à partir d'un comportement attendu, soit l'échelon C. Comme chaque commission scolaire, chaque école et parfois chaque enseignant peut choisir le type de notation qui lui convient, bien souvent en fonction des exigences du bulletin scolaire ou du type de notation qui avantage le plus ses élèves, il arrive que des élèves soient classés à partir de procédures différentes. En soi, cela ne pose pas de problème si les deux procédures produisent des classements identiques. En fait, ces dernières devraient produire des résultats similaires puisqu'il s'agit d'une épreuve standardisée.

Par ailleurs, le Ministère propose en parallèle à sa grille de correction détaillée en dix critères une grille dite abrégée en six critères : ces derniers incorporent les dix premiers dans une structure macroscopique.

Cette proposition vise vraisemblablement à atténuer le temps de correction imposé par la grille en dix points.

Aucune étude officielle ne permet actuellement de confirmer ou d'infirmer le bien-fondé de telles pratiques en regard de la validité intégrée à un processus de démarche évaluative où l'on doit tenir compte d'un ensemble de facettes et de préoccupations.

2. CADRE D'ANALYSE

L'examen de la pratique ministérielle en matière d'évaluation de la production écrite s'effectue dans le cadre du concept élargi de validité, intégré à une démarche évaluative et défini comme suit : «La validité est une évaluation globale, fondée sur des preuves empiriques et sur un rationnel théorique, de l'adéquation et de la justesse des inférences et des actions issues des scores de test» (Auger et Dassa, 1992).

La validité des tests telle qu'elle a été définie repose sur quatre points fondamentaux : a) une preuve de la convergence (ou de la divergence) des scores du test vers une signification du construit; b) une appréciation de la valeur des conséquences de l'interprétation des scores du test; c) une preuve de la pertinence du construit et de son utilité dans des situations spécifiques; d) une appréciation des conséquences sociales en relation avec les buts du *testing*. Pour mieux saisir toute la portée de cette définition, il faut souligner, avec Auger et Dassa, que :

> *Selon cette conception de la validité, la pertinence, l'interprétation et l'utilité des scores de test sont indissociables et doivent être intégrés au contexte de la démarche évaluative. Notons que les études empiriques portant sur la validité de construit ou de contenu touchent nécessairement à ces trois aspects. Ce qui signifie que tout test en éducation doit se référer à un domaine afin d'assurer une interprétation pertinente des résultats. Bien que ces deux formes de validité (de construit et de contenu) ne couvrent qu'une partie du concept de validité défini ci-haut, elles n'en constituent pas moins des composantes essentielles d'une conception intégrée de la validité.*

Quant à l'inférence dont il est question dans la définition précédente, elle dépend de la signification à accorder aux scores. Bien que cette signification repose sur la validité de contenu, la qualité de l'inférence est liée aussi à la qualité des preuves empiriques telles que la consistance des réponses ou bien la valeur de corrélations pertinentes.

Afin d'évaluer la capacité d'un test à mesurer ce qu'il doit mesurer, il convient d'élargir l'interprétation des scores de façon à tenir compte de leur utilisation. La validité d'un test porte alors sur l'évaluation de la valeur fonctionnelle du testing. Au-delà des données empiriques, la validité doit reposer sur l'appréciation des conséquences du testing et notamment l'évaluation des décisions qui en découlent. C'est pourquoi les scores du test doivent être mis en relation avec des construits pertinents de façon à pouvoir situer le test dans un contexte théorique de mesure. De même, l'interprétation des tests doit refléter les valeurs accordées aux comportements (désirables ou pas) que les scores expriment.

Le cumul des données, l'estimation de divers indices docimologiques de même que l'interprétation des données qualitatives constituent les assises des preuves présidant à l'interprétation des résultats obtenus aux tests. L'appréciation (exprimée par un jugement) des décisions qui découlent des tests établit les assises des conséquences de l'interprétation. L'évaluation de l'utilisation des tests implique l'étude de la pertinence et de l'utilité des tests et nécessite des jugements concernant leurs valeurs en termes de signification et de finalité. L'approche empirique reliée à la pertinence et à l'utilité des tests donne ses assises aux preuves relatives aux aspects de la validité propres à l'utilisation du test, tandis que l'approche par jugement confère des assises au regard des conséquences sociales de l'utilisation des tests (Auger et Dassa, 1992).

Cette intégration des assises (preuves et conséquences) aux fonctions (interprétation et utilisation) est résumée dans le tableau 1 ci-après. Le tableau original (Auger et Dassa, 1992) inclut autant l'aspect collectif

qu'individuel. Pour les besoins de cette contribution, seul l'aspect collectif est retenu. La conception de la validité des tests dans une démarche évaluative permet d'évaluer chaque étape (intention ou but, informations, jugement et décision) selon les assises et les fonctions précédentes.

Tableau 1 – La validité des tests dans une démarche évaluative en fonction d'un contexte d'évaluation sommative*

		Fonction	
		Interprétation	Utilisation
A S S I S E S	Preuves empiriques	Validité de contenu, de construit, concomitante, discriminante.	Validité reliée à un critère; prédictive; écologique. Fonctionnement différencié des items.
	Conséquences/ décisions	Règles de décision. Poids des composantes. Établissement de critères de passage ou score de césure. Établissement d'un intervalle de confiance autour des critères décisionnels. Choix des juges et protocole de compétence. Adéquation globale au profil attendu de compétence scolaire.	Pertinence et cohérence de la méthodologie utilisée (incluant les procédures de transformation des notes). Pertinence et cohérence des décisions concernant les populations scolaires en regard des finalités du testing.

* Source – Auger et Dassa (1992).

En regard des buts et du type de décision visés par le Ministère, les formes pertinentes de validité devraient permettre un passage à la classe supérieure de qualité. Cela signifie qu'au niveau de l'interprétation, il conviendrait de s'assurer qu'un jugement de maîtrise suffisante des apprentissages (Auger, 1991) s'appuie sur un ou des instruments qui possèdent à tout le moins une validité de contenu et une validité discriminante adéquates. Un examen des conséquences de la fonction d'interprétation du tableau 1 fait ressortir la nécessité d'une appréciation globale du profil attendu de compétence scolaire lors de l'élaboration de règles de décision et de l'établissement de critères de passage ou de scores de césure. Une telle nécessité conférerait de meilleures assises à la validité discriminante.

En ce qui concerne l'utilisation de l'épreuve ministérielle, un examen de la performance des élèves au regard des thèmes proposés (un choix parmi trois) ne devrait démontrer aucun fonctionnement différencié selon le thème choisi. De plus, une vérification de la capacité de l'épreuve à prédire la réussite scolaire en première secondaire serait fort indiqué. Un examen de l'équivalence des transformations des notes (avec ou sans valeur numérique) en cinq catégories (très bonne maîtrise, bonne maîtrise, maîtrise suffisante, maîtrise insuffisante, échec) et de l'abrégement de la grille de correction à six critères pour assurer une validité écologique informerait sur la pertinence et sur l'utilité de la méthodologie employée en fonction des décisions recherchées. Rappelons que la validité écologique porte sur «le degré de ressemblance entre les conditions dans lesquelles se trouvent les sujets et celles de la population entière» (Legendre, 1988, 635). Entendons ici les conditions dans lesquelles les sujets sont contraints et celles qui sont vécues habituellement durant l'année. L'ensemble de ces analyses apporterait des preuves empiriques et une appréciation du bien-fondé de l'utilisation de l'épreuve pour le but poursuivi.

3. MÉTHODOLOGIE

Pour procéder à l'étude de la validité sous ses multiples facettes, deux épreuves de production écrite sont choisies, celle de juin 1990 et celle de juin 1991. En juin 1990, il s'agissait pour l'élève d'écrire un

texte en vue d'inciter les jeunes de son âge à agir à propos de la natation, de la lecture ou des emplois d'été. Par ailleurs, en juin 1991, les élèves devaient composer un texte d'opinion soit sur leur place dans la famille, soit sur l'intégration au groupe quand on est différent, soit sur le choix d'une carrière par goût ou pour l'argent. Pour un aperçu de la grille de correction et des interprétations inférées à partir de valeurs numériques ou non, voir l'annexe 1. Pour chacune des épreuves, un échantillon provincial (cinquante écoles au hasard) totalisant six cents copies d'élèves est à notre disposition. Les résultats proviennent d'une correction, centralisée et supervisée par une professionnelle du Ministère.

L'équipe de correction, constituée de sept enseignants de sixième année reconnus pour leurs compétences et pour leur expérience dans la production écrite de sixième année, a reçu une formation de deux jours. Cette formation comprenait : a) une information générale sur l'épreuve et, en particulier, sur chacun des critères de correction; b) une correction individuelle d'un même texte suivie d'une mise en commun; c) une discussion quant à chacun des critères lorsqu'il existait des écarts. Ces deux dernières étapes étaient reprises plusieurs fois. Malgré une procédure systématique, aucun indice de fidélité intercorrecteurs n'est disponible.

Les résultats à ces deux épreuves servent à l'étude de la structure des grilles de correction par le biais d'analyses factorielles et principalement à l'étude de la validité discriminante entre le classement des élèves en cinq catégories selon l'échelle sans valeur numérique (critère dichotomique et proposition de cinq catégories de maîtrise selon le nombre de critères réussis sur dix) et l'échelle avec valeurs numériques (note pondérée sur cent). Dans les deux cas, les dix critères de correction sont utilisés. L'analyse discriminante permet ainsi de vérifier l'équivalence du classement des élèves selon les échelles avec ou sans valeur numérique. Aucune donnée n'est disponible pour l'examen des conséquences de la grille abrégée sur le classement des élèves. Néanmoins, l'analyse discriminante permet de suggérer quelques pistes de réflexion en ce qui concerne la pondération des critères en vue d'une optimisation du classement des élèves et l'analyse factorielle permet de suggérer une structure plus globale des regroupements des critères.

Les mêmes résultats de base servent également à l'étude des résultats globaux des élèves par rapport aux trois thèmes proposés pour les épreuves écrites. Cette comparaison s'effectue par une analyse de la variance où la variable indépendante est constituée des trois thèmes et où la dépendante correspond au score total (pondéré selon l'importance de chacun des critères) obtenu sur l'échelle numérique.

La validité est discutée à partir des propositions au tableau croisé entre les assises (preuves empiriques et conséquences/décisions) et les fonctions (interprétation et utilisation). Seules les propositions adéquates au regard des buts, des types de décision et selon les données disponibles sont retenues pour les fins de la présente analyse. De fait, pour le croisement preuves empiriques et interprétation, la validité de contenu, de construit et discriminante sont retenues; pour le croisement conséquences/décisions et interprétation, l'appréciation de l'adéquation du seuil de réussite comme critère décisionnel et l'apport de l'importance relative de chacun des critères de correction sont étudiés; pour le croisement preuves empiriques et utilisation, l'analyse du fonctionnement différencié des thèmes apporte des preuves afin d'appuyer la pertinence et l'utilité de proposer un choix de thèmes; pour le croisement conséquences/décisions et utilisation, l'étude de l'équivalence des transformations des notes sur les deux échelles permet d'apprécier cette modalité au regard de sa contribution au but visé.

4. ANALYSE

La validité de contenu des épreuves fut assurée par des jugements obtenus à la suite des réunions de spécialistes en français sur le plan provincial; en outre, la grille de correction respecte dans l'ensemble les critères docimologiques. Cette étape s'insère dans la démarche d'élaboration des épreuves ministérielles.

La validité de construit est examinée par le biais d'analyses factorielles. Ces dernières semblent démontrer le bien-fondé de la structure générale de la grille de correction. En effet, le Ministère présente l'ensemble des critères en cinq groupements distincts : choix des informations (critères 1, 2), organisation du texte (critères 3, 4, 5), organisation de la phrase ((critères 6, 7), respect du code lexical (critère 8), respect

du code grammatical (critères 9, 10). Pour l'épreuve de juin 1990, les méthodes de factorisation «alpha» et du «maximum de vraisemblance» avec rotation «oblimin» produisent des solutions satisfaisantes à quatre facteurs et expliquent 49 % de la variance commune. Pour l'épreuve de juin 1991, les mêmes procédures donnent des solutions à quatre facteurs et expliquent 42 % de la variance commune. Des résultats plus détaillés de ces analyses se retrouvent à l'annexe 2. À l'épreuve de juin 1990 (texte incitatif), on retrouve presque les mêmes regroupements que ceux de la grille de correction, à l'exception des critères 6 et 7. Ce dernier constitue à lui seul un facteur distinct. Pour l'épreuve de juin 1991 (texte d'opinion), on observe : a) un regroupement identique en ce qui concerne le respect des codes lexical et grammatical (critères 8,9,10), et la ponctuation (critère 6); b) un déplacement du critère 3 (arguments regroupés logiquement) vers les critères reliés au choix des informations; c) un regroupement des critères 4 (liens appropriés et référents) et 7 (phrases simples et composées bien construites); d) l'isolement du critère 5 (découpage en paragraphes). Pour les deux épreuves, ce sont les critères reliés à l'organisation de la phrase (critères 6, 7) qui constituent une structure plus mobile. Par contre, le critère 6 rejoint les critères 8, 9, 10 et forme un regroupement stable et souhaitable. L'écart entre ce qui est observé par les analyses factorielles et la structure actuelle des épreuves pourrait s'expliquer : a) par des définitions opérationnelles qui s'éloignent du construit recherché et par une correction qui associe de trop près les critères reliés à la construction de phrases et les liens ou référents; b) par une structure légèrement différente en fonction des types de discours et des thèmes utilisés. La prudence s'impose donc quant à des conclusions trop hâtives puisque les analyses factorielles précédentes ont été produites uniquement dans une perspective exploratoire.

Tel que cela a été précisé précédemment, on ne peut examiner directement la proposition ministérielle au sujet de l'abrégement des dix critères de correction en six critères macroscopiques. En revanche, les résultats des analyses factorielles suggèrent quatre regroupements possibles. À partir de ces résultats, il est plausible de penser à abréger la grille détaillée en quatre critères macroscopiques.

Les interprétations du degré de maîtrise de l'habileté à écrire proviennent de l'étude du profil de maîtrise suffisante des élèves de sixième année et des diverses opérations d'établissement de seuils de réussite effectuées au cours des années antérieures (Auger, Brunelle et Ouellet, 1989). C'est à partir de cette étude que la catégorie de maîtrise suffisante fut établie à huit critères réussis sur 10. Ce seuil de réussite est le résultat de l'application d'une variante de la méthode d'Angoff (1971) associé à une animation en convergence (Auger, 1989). Au cours de la démarche, les juges experts doivent apprécier chacun des critères et les rendre dichotomiques avant de passer à l'établissement proprement dit du seuil de réussite. Des études de généralisabilité (Cardinet et Tourneur, 1985) effectuées à l'aide du logiciel ETUDGEN (1987) donnent des coefficients d'assurance absolue ou des coefficients de fidélité interjuges autour de 0,90. L'ensemble des quatre autres catégories (très bonne maîtrise, bonne maîtrise, maîtrise insuffisante et échec) constitue une proposition ministérielle fait à la suite des analyses descriptives.

Les analyses relatives à la validité discriminante des épreuves de juin 1990 et de juin 1991, en ce qui a trait à l'équivalence des classements des élèves à partir des échelles avec et sans valeur numérique donnent respectivement des proportions correctes de classification de 72,83 % et de 75,50 %. Le classement des élèves est ici considéré selon cinq degrés de maîtrise tels qu'ils sont fixés par le profil de compétence attendue et les variables discriminantes constituées des critères pondérés de correction sur l'échelle numérique. Les résultats obtenus proviennent des contraintes imposées suivantes :

- la variable classement en cinq degrés de maîtrise sans valeur numérique est de type ordinal;
- les variables discriminantes (10 critères de correction évalués en fonction d'une cote de base (1 à 5) constituent une échelle de mesure à intervalles égaux;
- chacune des variables discriminantes n'est pas considérée comme une combinaison linéaire d'une autre, c'est-à-dire qu'aucune des variables n'est redondante en termes d'information;

- les étapes dans la sélection des variables discriminantes ont suivi la règle de la valeur minimale du lambda de Wilk et d'un niveau de tolérance minimum de 0,001;
- la probabilité de classification tient compte d'une information *a priori* donnée par la distribution empirique des élèves.

Le tableau 2 détaille le classement des élèves pour chacune des catégories de maîtrise prise deux à deux et permet de comprendre les taux de classification moyens pour l'ensemble des catégories. Le taux de classification le plus bas correspond à la prise de décision la plus importante soit celle qui distingue les élèves qui ont une maîtrise suffisante de ceux qui ont une maîtrise insuffisante. On observe un bon taux de classification pour les catégories extrêmes, soit autour de 90 %. Il est à noter que pour les deux épreuves les taux de classification sont similaires même si le type de texte est différent.

Tableau 2 – Taux de classification corrects pour chacune des catégories de maîtrise prise deux à deux et pour l'ensemble des cinq catégories

Catégorie	Juin 1990	Juin 1991
Très bonne maîtrise/ bonne maîtrise	92,4 %	90,6 %
Bonne maîtrise/ maîtrise suffisante	83,5 %	87,2 %
Maîtrise suffisante/ maîtrise insuffisante	78,6 %	78,8 %
Maîtrise insuffisante/ échec	89,6 %	90,1 %
Très bonne maîtrise à Échec	*72,83 %*	*75,50 %*

Une analyse complémentaire en regard de la classification des élèves en deux macrocatégories (les élèves qui maîtrisent et ceux qui ne maîtrisent pas) donne un taux de classification de l'ordre de 90,5 % pour l'épreuve de juin 1991 et de 92 % pour juin 1990. Cela justifie l'établissement du seuil de réussite à 8 critères sur 10 et son équivalent sur l'échelle avec valeur numérique, soit une note de 62 sur 100.

En somme, l'ensemble des analyses discriminantes fait ressortir le bien-fondé du classement à partir du seuil de réussite et la fragilité de l'interprétation autour des quatre autres catégories de maîtrise pour les échelles avec ou sans valeur numérique.

La pondération des critères suggérée à la suite des analyses discriminantes offre quelques pistes de réflexions pour fixer des poids relatifs qui pourraient être plus valables par rapport au croisement conséquences/décisions et interprétation que celles qui sont proposées actuellement. Le point de vue inhérent à l'analyse discriminante est celui de la pondération en vue d'une meilleure prise de décision. Pour les cinq catégories de maîtrise, l'analyse estime des coefficients standardisés à la fonction discriminante. Ces coefficients permettent de proposer un poids relatif à chacun des critères de correction en vue d'optimiser le taux de classification. Pour l'épreuve de juin 1990, la première fonction discriminante explique 97 % de la variance totale et correspond à un coefficient de corrélation canonique de 0,90. Cette corrélation est une mesure d'association entre la fonction discriminante et la classification entre les groupes de maîtrisants. À l'épreuve de juin 1991, la première fonction discriminante explique 95 % de la variance totale et possède un coefficient de corrélation canonique de 0,89.

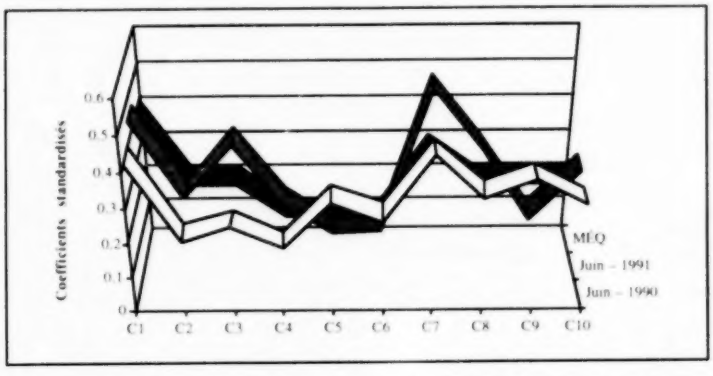
En fonction de ces constats, la figure 1 représente l'importance relative à accorder aux critères de correction pour les deux épreuves respectives en considérant le classement dans les cinq catégories de maîtrise. Pour juin 1990, les critères dans l'ordre d'importance sont les 7, 1, 9, 8, 5, 10, 6, 3, 2, 4 et, pour juin 1991, les 7, 1, 3, 8, 10, 2, 4, 9, 6, 5. Les deux premiers critères sont les mêmes, soit les suivants : les phrases simples et composées sont bien construites et les textes comportent des informations pertinentes qui soutiennent l'opinion. Pour le texte d'opinion (épreuve 1991), le critère 3 relatif aux informations groupées logiquement prend plus d'importance, tandis que le critère 8 relié au respect du code lexical est stable en quatrième position. Le critère 10 qui concerne la conjugaison est relativement stable.

Le Ministère s'appuie sur une importance relative des critères en fonction du contenu et non en fonction d'un pouvoir de discrimination.

Ainsi, il propose une pondération plus élevée pour le choix des informations (critère 10) selon un rapport de 4:1 et la structure des phrases (critère 7) de 3:1, une importance moyenne aux critères 2, 3, 8, 9, 10 de 2:1, un poids relativement peu important pour les critères 4, 5, 6 de 1:1.

Selon l'analyse des résultats basée sur les deux épreuves et considérant les coefficients standardisés à la fonction discriminante, il y aurait une tendance à accorder une importance relative plus élevée pour les critères 1 et 7 (4:1, en juin 1990 et 5:1 en 1991), à attribuer une importance moyenne pour les critères 8 et 10 (3:1) et à attribuer une importance variable selon le type de texte pour les autres critères.

Les analyses font également ressortir la possibilité de pondérer les critères de correction différemment selon les types de discours.



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Juin 1990	0,41	0,19	0,23	0,17	0,30	0,25	0,43	0,31	0,36	0,29
Juin 1991	0,46	0,24	0,39	0,22	0,13	0,14	0,54	0,37	0,17	0,31
Ministère	0,40	0,20	0,20	0,10	0,10	0,10	0,30	0,20	0,20	0,20

Figure 1 – Pondération des critères selon la première fonction discriminante pour l'ensemble des catégories de maîtrise aux deux épreuves de production écrite et pondération proposée par le Ministère

L'étude du fonctionnement différencié des thèmes par l'analyse de la variance de la performance des élèves par sujet à l'épreuve de juin 1991 est statistiquement significative à une probabilité de 5 % tandis qu'à l'épreuve de juin 1990, le test F est non significatif. Le test de l'égalité des variances est non significatif pour les deux épreuves. Il ressort que les moyennes des élèves à l'épreuve de juin 1991 sont différentes selon les thèmes. Il existe donc une différence dans le fonctionnement général de l'épreuve selon le thème choisi. On est donc à même de questionner la pertinence de proposer plus d'un choix de sujet, surtout lorsqu'il s'agit d'une épreuve standardisée et servant pour la promotion des élèves. D'autant plus qu'aucune étude ne vient confirmer la fidélité intercorrecteurs, et ceci, par sujet proposé.

CONCLUSION

L'étude globale de la validité de la pratique ministérielle en matière d'évaluation sommative de la production écrite en français s'est effectuée à partir d'un cadre d'analyse qui propose un concept de la validité élargi et intégré à des étapes d'une démarche évaluative. Cette dernière doit assurer une cohérence entre le but visé par l'épreuve, l'instrument de mesure et ses modalités d'utilisation, les interprétations reliées aux scores et les décisions à prendre. L'examen de la validité sous ses multiples facettes a fait ressortir les conclusions suivantes, pour les croisements qui suivent.

Preuves empiriques et interprétation. – La validité du contenu et la règle de décision reliée au seuil de réussite sont justifiées; la grille de correction en production écrite respecte globalement une structure empirique qui serait vraisemblablement conforme à une validité de construit, quoique certaines améliorations soient souhaitables, particulièrement pour les critères 4, 6 et 7.

Conséquences/décisions et interprétation. – L'établissement du seuil de réussite à huit critères sur dix se justifie par l'application rigoureuse de la méthode utilisée et par un degré de classification de l'ordre de 90 % entre les élèves qui maîtrisent et ceux qui ne maîtrisent pas, et selon les échelles avec ou sans valeur numérique. La pondération autour des critères de correction devrait être différente afin d'optimiser le

taux de classification des élèves en cinq catégories de maîtrise et ce, pour chacun des deux types de discours.

Preuves empiriques et utilisation. Le fait de donner le choix des thèmes d'écriture aux élèves peut favoriser un fonctionnement différencié de leurs performances.

Conséquences/décisions et utilisation. – L'équivalence de la classification des élèves en cinq catégories de maîtrise sur les deux échelles (avec et sans valeur numérique) n'est pas suffisamment prouvée et l'on ne peut pas justifier son utilisation pour des épreuves standardisées.

La pondération des critères de correction pose le dilemme de la détermination des poids relatifs selon une préoccupation de testing ou selon l'importance que l'on accorde au contenu au regard de l'apprentissage de l'habileté à écrire. C'est ici que l'on pourrait préférer la validité écologique au détriment de la validité discriminante. La décision devrait être prise en fonction du but principalement visé par la démarche évaluative, soit un meilleur classement des élèves et tenant compte dans la mesure du possible des types de discours.

Le fait de laisser le choix du thème aux élèves afin de favoriser une plus grande possibilité de produire un texte à la hauteur de leur capacité se défend en soi. Mais du point de vue du testing et au regard du but visé par une telle opération, cela laisse plus d'interrogations que de certitudes. Entre autres, il y aurait lieu de penser à ne proposer qu'un seul thème par épreuve.

La proposition des classements en cinq catégories de maîtrise sur deux échelles différentes ne semble pas viable actuellement, sauf pour les macrocatégories (maîtrisants et non-maîtrisants). Si l'on croit à une interprétation critérielle basée sur des comportements attendus, on aurait avantage à procéder à une analyse pour distinguer plus clairement chacune des catégories proposées autant par l'apport de preuves empiriques que par l'examen des conséquences sur les décisions. Ceci permettrait d'améliorer la pertinence par rapport à l'utilisation des épreuves pour les besoins de la promotion des élèves. Le Ministère devrait pou-

voir choisir un type d'échelle ou proposer un classement uniquement à partir des deux grandes catégories.

Évidemment, le tout repose sur le constat que malgré l'absence d'un coefficient de fidélité intercorrecteurs, les données fournies pour les fins de l'étude proviennent d'une procédure systématique de correction. Les données disponibles ne permettent pas de conclure au sujet de la grille de correction abrégée. C'est pourquoi la proposition de l'abrégement des critères demande à être examinée à partir d'autres données réelles pour consolider les assises en fonction de preuves empiriques et de l'évaluation des conséquences au regard des décisions. De plus, un examen de la validité prédictive permettrait de mieux fonder le concept de maîtrise suffisante en apportant des preuves empiriques supplémentaires des compétences scolaires évaluées.

En général, en examinant les pratiques évaluatives ministérielles québécoises en production écrite dans le cadre de la validité des tests intégrée à une démarche évaluative, les analyses précédentes ont mis en relief le fait que les pratiques usuelles, en se limitant à certaines composantes de la validité (de contenu, étude sur l'établissement de seuils de réussite), n'atteignent qu'en partie le but visé par l'instrument de mesure au regard de l'interprétation et de l'utilisation des scores aux épreuves. De ce fait, ces pratiques sont parcellaires, car elles intègrent difficilement le processus d'évaluation fondée sur le paradigme information-jugement-décision (Auger et Dassa, 1992).

RÉFÉRENCES

AUGER, R. (1992).
Enseigner pour évaluer? – Évaluer pour un meilleur enseignement? Évaluer pour un meilleur réinvestissement? Évaluer pour vérifier ses hypothèses quant aux causes possibles? Communication verbale et table ronde dans le cadre de l'Association québécoise de professeurs de français, section Montréal, Château Champlain.

AUGER, R. et DASSA, C. (1992).
Les pratiques de mesure et d'évaluation des apprentissages et la validité des tests dans un contexte de démarche évaluative. In Laveault, D. (éd.), *Les pratiques d'évaluation en éducation* (p. 151-166). Hull : ADMÉE.

AUGER, R. (1991).
De la définition des concepts à la pratique – Réflexions et suggestions. Critique et communication écrite de M.R. Jeager, The psychometric demands of testing of licensure and certification. Dans le cadre du symposium international des Modern Theories in Measurement : Issues and Practices. Québec : Le Château Montebello.

AUGER, R. (1989).
Étude de praticabilité du testing adaptatif de maîtrise en regard des apprentissages scolaires au Québec – Une expérimentation en éducation économique secondaire 5. Thèse de doctorat, Université du Québec à Montréal.

AUGER, R., BRUNELLE, M. et OUELLET, L. (1989).
Établir un seuil de réussite : une nécessité en évaluation sommative ? Session d'études au colloque international de l'ADMÉE.

ANGOFF, W. (1971).
Scales, norms and equivalent scores. *In* Thorndike, R.L. (éd.), *Educational measurement* (2e édition) (p. 508-600). Washington (DC) : ACE.

CARDINET, J. et TOURNEUR, Y. (1985).
Assurer la mesure, guide pour les études de généralisabilité. Berne : Peter Lang.

ETUDGEN (1987).
Logiciel d'analyse de la variance pour la théorie de la généralisabilité, version modifiée pour IBM compatible par Bertrand, R. et Dufour, N. Université Laval, Québec, d'une version précédente de Duquesne, F., Université de Mons, Belgique.

LEGENDRE, R. (1988).
Dictionnaire actuel de l'éducation. Paris : Larousse.

MINISTÈRE DE L'ÉDUCATION (1986).
Document d'information. Une épreuve ministérielle de production écrite en 6e année du primaire et en 5e année du secondaire. Québec : Direction générale de l'évaluation.

ANNEXE I

INFORMATIONS AU SUJET DE LA GRILLE DE CORRECTION ÉCRITURE 6e, JUIN 1991, MINISTÈRE DE L'ÉDUCATION DU QUÉBEC

a) Interprétation à l'aide de la grille de correction sans valeur numérique

Seuil de réussite proposé : huit éléments réussis sur 10
(un élément est réussi si l'élève obtient la cote A, B ou C).

Résultat de la mesure	Interprétation	Jugement	Cote
10 éléments réussis sur 10 dont les éléments : 1 (choix des informations); 3 (cohérence) et 7 (structure des phrases ont la cote A)	Très au-dessus du seuil de réussite	Très bonne maîtrise	M++
9 ou 10 éléments réussis sur 10	Au-dessus du seuil de réussite	Bonne maîtrise	M+
8 éléments réussis sur 10	Égal au seuil de réussite	Maîtrise suffisante	M
6 ou 7 éléments réussis sur 10	Au-dessous du seuil de réussite	Maîtrise insuffisante	M-
5 éléments réussis et moins	Très au-dessous du seuil de réussite	Échec	E

b) Interprétation à l'aide de la grille de correction avec valeur numérique

Seuil de réussite proposé : 62 sur 100.

Résultat de la mesure	Interprétation	Jugement	Cote
De 91 à 100	Très au-dessus du seuil de réussite	Très bonne maîtrise	M++
De 71 à 90	Au-dessus du seuil de réussite	Bonne maîtrise	M+
De 62 à 70	Égal au seuil de réussite	Maîtrise suffisante	M
De 53 à 61	Au-dessous du seuil de réussite	Maîtrise insuffisante	M-
Moins de 52	Très au-dessous du seuil de réussite	Échec	E

Source – Guide d'administration et de correction. Extrait de l'épreuve 014-600-00, p. 23.

INFORMATIONS AU SUJET DE LA GRILLE DE CORRECTION ÉCRITURE 6e JUIN 1991, MINISTÈRE DE L'ÉDUCATION DU QUÉBEC

c) Grille de correction détaillée

Critère	Élément	Échelle descriptive				
		A	B	C	D	E
Choix	Le texte comporte des informations pertinentes qui soutiennent l'opinion.	Informations très pertinentes. Nombre suffisant ou plus suffisant.	Informations pertinentes Nombre suffisant ou plus que suffisant	Informations pertinentes Nombre suffisant	Informations pertinentes Nombre suffisant	Texte hors du sujet
	Le vocabulaire est adapté au projet : il est varié et correct.	Vocabulaire riche et très bien adapté au projet	Vocabulaire riche et bien adapté au projet	Vocabulaire simple et adapté au projet	Vocabulaire pauvre et peu adapté au projet	Vocabulaire très pauvre et peu adapté au projet
Organisation du texte	Les informations sont groupées logiquement.	Texte très cohérent Toutes les informations sont groupées	Texte cohérent Une information déplacée	Texte cohérent Deux informations déplacées	Texte peu cohérent Trois ou quatre informations déplacées	Texte incohérent Informations très éparpillées
	Les informations entre les phrases sont liées.	Liens presque toujours appropriés Aucune faute	Liens très souvent appropriés Une ou deux fautes	Liens assez souvent appropriés Trois ou quatre fautes	Liens souvent impropres Cinq ou six fautes	Liens très souvent impropres Sept fautes et plus
	Le texte est bien découpé en paragraphes.	Les paragraphes correspondent parfaitement aux idées développées.	Les paragraphes correspondent très bien aux idées développées.	Les paragraphes correspondent globalement aux idées développées	Emploi abusif de paragraphes ou faiblesse dans l'introduction et la conclusion	Aucun découpage

c) Grille de correction détaillée (suite)

Organisation	Les signes de ponctuation sont utilisés adéquatement.	Moins de 1 % de fautes	1 % de fautes	2 % de fautes	9 % de fautes	Plus de 3 % de fautes
de la phrase	Les phrases, simples et composées, sont bien construites.	Toutes les phrases ont du sens et sont bien construites. C.* 0 faute M.* 0 faute L.* 0-1 faute	Presque toutes les phrases ont du sens et sont bien construites. C. 1 faute M. 1 faute L. 2 fautes	La plupart des phrases ont du sens et sont bien construites. C. 2 fautes M. 2-3 fautes L. 3-4 fautes	Plusieurs phrases comportent des fautes de sens ou de construction C. 3-5 fautes M. 4-6 fautes L. 5-7 fautes	De nombreuses phrases comportent des fautes de sens ou de construction C. 6 et plus M. 7 et plus L. 8 et plus
Respect du code lexical	Les mots sont écrits correctement.	Moins de 2 % de fautes	2 % de fautes	3 % de fautes	4 % de fautes	Plus de 4 % de fautes
Respect du code grammatical	Les norms et les adjectifs sont bien accordés.	Moins de 1 % de fautes	1 % de fautes	2 % de fautes	3 % de fautes	Plus de 3 % de fautes
	Les finales des verbes sont correctes.	Moins de 1 % de fautes	1 % de fautes	2 % de fautes	3 % de fautes	Plus de 3 % de fautes

Source – Guide d'administration et de correction. Extrait de l'épreuve 014-600-00, p. 23.

* C : texte court (moins de 250 mots)

M : texte moyen (250 à 349 mots)

L : texte long (350 mots et plus)

ANNEXE 2

ANALYSES FACTORIELLES POUR LES ÉPREUVES MINISTÉRIELLES DE JUIN 1990 ET DE JUIN 1991 EN PRODUCTION ÉCRITE

Relevés sommaires

Juin 1990

Variable	Communality*		F	Loadings	Pct of Var	Cum Pct
C1	0,56223	*	1	3,57266	35,7	35,7
C2	0,57736	*	2	0,90041	9,0	44,7
C3	0,55051	*	3	0,21521	2,2	46,9
C4	0,56356	*	4	0,19344	1,9	48,8
C5	0,45245	*				
C6	0,32721	*				
C7	0,27573	*				
C8	0,65433	*				
C9	0,46536	*				
C10	0,45299	*				

Analyses factorielles

Oblimin Rotation 1, Analysis 1 - Kaiser Normalization.
Oblimin converged in 39 iterations.

	F1	F2	F3	F4	
C5	0,55349	-0,09066	0,18895	0,15397	
C4	0,54762	-0,08593	-0,27296	0,04038	Organisation du texte
C3	0,48203	0,02166	-0,16367	0,25772	
C8	-0,01512	-0,82502	0,07127	0,02442	Respect du code lexical
C9	-0,10738	-0,68669	0,00150	0,07657	Respect du code grammatical
C10	0,16164	-0,50058	-0,04922	-0,06290	Ponctuation
C7	0,00436	-0,12931	-0,35463	0,17283	Organisation de la phrase
C1	0,02382	-0,03472	0,02491	0,72553	Choix des informations
C2	0,13113	-0,00578	-0,13067	0,60222	

Juin 1991

Variable	Communality*		F	Loadings	Pct of Var	Cum Pct
C1	0,39055	*	1	2,93497	29,3	29,3
C2	0,33960	*	2	0,74329	7,4	36,8
C3	0,54845	*	3	0,30168	3,0	39,8
C4	0,43841	*	4	0,18708	11,9	41,7
C5	0,29811	*				
C6	0,28457	*				
C7	0,42838	*				
C8	0,59542	*				
C9	0,30013	*				
C10	0,54340	*				

Analyses factorielles

Oblimin Rotation 1, Analysis 1 - Kaiser Normalization.
Oblimin converged in 19 iterations.

	F1	F2	F3	F4	
C8	0,79449	-0,05510	-0,09755	0,01172	Respect du code lexical
C10	0,74352	0,00510	0,03090	0,13466	Respect du code grammatical
C9	0,49424	-0,03162	0,03230	-0,06655	
C6	0,44322	0,05911	0,12931	-0,12028	Ponctuation
C1	0,00972	-0,65407	-0,06741	0,01371	Choix des informations
C3	-0,05514	-0,64153	0,22071	0,00834	Organisation du texte
C2	0,06288	-0,51415	0,05369	-0,02243	Choix des informations
C7	0,11819	-0,11052	0,54179	0,14871	Organisation de la phrase
C4	0,05259	-0,06953	0,49129	-0,25481	Organisation du texte
C5	0,09352	-0,30344	-0,01648	-0,31962	Organisation du texte

OUVRAGES DISPONIBLES AUX ÉDITIONS DU CRP

ADAPTATION SCOLAIRE

- *Classes sociales et motricité*. D. HADE (1989).
- *L'intégration d'enfants handicapés*. J. GIGNAC (1988).

ADMINISTRATION SCOLAIRE

- *L'évaluation institutionnelle en milieu scolaire – Logiques, enjeux, rôles et responsabilités des différents acteurs*. D. MASSE (dir.) (1993).
- *Pratiques et modèles de formation en administration scolaire*. A. GODIN (dir.) (1993).

ANDRAGOGIE

- *Formation-travail / Travail-formation* (1er vol. – *Formation, travail et savoir institué*; 2e vol. – *Pratiques de développement et de reconnaissance des compétences*). J.-L. LEVESQUE, J. FERNANDEZ et M. CHAPUT (dir.) (1993).
- *Recherche, formation et pratiques en éducation des adultes*. F. SERRE (dir.) (1993).

DIDACTIQUE

- *Didactique du français langue maternelle – Bilan des recherches québécoises*. G. GAGNÉ et G.-R. ROY (1989).
- *Genre et nombre* (1er vol. – *Étude théorique*; 2e vol. – *Étude appliquée*). G.-R. ROY (1985).
- *La pédagogie du brouillon*. L. MCNICOLL et G.-R. ROY (1989).
- *Le Fabloc* (maître et élève). S. POULIOT (1986).
- *L'enseignement des sciences humaines au primaire*. Y. LENOIR et M. LAFOREST (dir.) (1991).
- *L'interdisciplinarité : une voie d'avenir?* R. DELISLE et P. BÉGIN (dir.) (1992).
- *Lire et rédiger au bureau* (maître). C. GAUDREAU (1990).
- *Lire et rédiger au bureau* (élève). C. GAUDREAU (1990).
- *Manuel d'apprentissage de l'opérateur et du programmeur de MOCN*. (4 cahiers). M. BRÉZINA et DENIS DUBÉ (1985).
- *Musique et psaltérion* (guide et volume de pièces harmonisées). F. CORPATAUX (1987; édition révisée, 1992).
- *Phénomènes électriques en soudage*. (5 cahiers). P.-A. SIMARD, R. PARENT et H. HENSLER (1989).
- *Phénomènes chimiques en soudage*. (2 cahiers). P.-A. SIMARD, R. PARENT et H. HENSLER (1990).
- *Recherche et formation technologique*. P.-A. SIMARD et G. ROBIDAS (éd.) (1988).
- *Réussir dès l'entrée dans l'écrit*. G. BOUDREAU (1993).
- *S'approprier l'orthographe grammaticale par l'approche «donneur → receveur»*. G.-R. ROY et H. BIRON (1991).
- *Sens des didactiques et didactique du sens*. P. JONNAERT et Y. LENOIR (dir.) (1993).

OUVRAGES DISPONIBLES AUX ÉDITIONS DU CRP

DIDACTIQUE (suite)

- *Étude de la maîtrise du français écrit à l'université.* G.-R. ROY et L. LAFONTAINE (1992).
- *Vers un triple regard sur le français écrit des étudiants de collèges et d'universités.* G.-R. ROY, L. LAFONTAINE, G. BOUDREAU et R. VIAU (1992).

INTERCULTUREL

- *Éducation et pédagogie interculturelles.* A. BEAUCHESNE (1991).
- *Les agents de liaison de Montréal, Toronto et Vancouver.* A. BEAUCHESNE (1987).

ORIENTATION PROFESSIONNELLE

- *Explorons des professions techniques et scientifiques.* P. DUPONT, S. MARCEAU et M. GINGRAS (1988).
- *La formation et l'insertion professionnelle, enjeux dominants dans la société postindustrielle.* C. LAFLAMME (dir.) (1993).
- *Future Between us – The Participants' Manual.* J. LIMOGES (1992)
  Training Peer Career Counsellors – The Trainer's Manual. J. LIMOGES (1992)

PÉDAGOGIE

- *La planification de l'enseignement – Deux approches, deux visions?* R. VIAU (dir.) (1993).
- *La recherche en formation des maîtres : détour ou passage obligé sur la voie de la professionnalisation?* H. HENSLER (dir.) (1993).
- *Le petit groupe d'apprentissage comme formule pédagogique.* L. PRONOVOST et A. BEAUCHESNE (1991).
- *L'évaluation des apprentissages : réflexions, nouvelles tendances et formation.* R. HIVON (dir.) (1993).
- *Vivre à la maternelle.* R. FORTIER (1988).

PSYCHOLOGIE

- *Développement en tête!* J. LIMOGES et R. P. HÉBERT (1988).
- *Le développement du moi.* J. LIMOGES et D. PAUL (1990).

DIVERS

- *Contenus et impacts de la recherche universitaire actuelle en sciences de l'éducation.* Actes du 2e congrès des sciences de l'éducation de langue française du Canada. (3 tomes). ROY, G.-R. (dir.) (1990).
- *Répertoire des publications 1988 des professeurs en sciences de l'éducation des universités du Québec.* (1990).
- *Répertoire des publications 1989 et 1990 des professeures et des professeurs des départements et facultés des sciences de l'éducation des universités du Québec.* (1991).

L'évaluation des apprentissages est discutée et souvent très contestée par ceux qui la vivent, élèves, étudiants et formateurs. Elle est une préoccupation majeure des formateurs, des administrateurs et des personnes évaluées. Elle est souvent remise en question.

La problématique de l'évaluation demeure vive : celle de la vérification des compétences, celle de l'aide à l'apprentissage, celle de la responsabilité professionnelle, celle du dilemme de sa validité et du jugement personnel.

Les recherches récentes en mesure et évaluation, en psychologie cognitive, celles relatives aux pratiques et aux croyances des enseignants, l'informatisation sont autant de sources nouvelles de réflexions et de pratiques contrôlées. Aux concepts spécifiques de validité et d'évaluation sommative, d'autres sont mis en relation : régulation, métacognition, objectifs d'intégration et évaluation authentique.

Ont collaboré à la réalisation de cet ouvrage

Linda ALLAL, Université de Genève
Réjean AUGER, Université du Québec à Montréal
Louise BÉLAIR, Université d'Ottawa
Étienne BOURGEOIS, Université catholique de Louvain
Rose BURELLE, Université du Québec à Montréal
Clément DASSA, Université de Montréal
Jean-Marie De KETELE, Université catholique de Louvain
René HIVON, Université de Sherbrooke
Michel LAURIER, Université de Montréal
Carmen PARENT, Université du Québec à Montréal
Richard ROBILLARD, Université Sainte-Anne
Serge P. SÉGUIN, Université du Québec à Montréal
Jacques TARDIF, Université de Sherbrooke

ISBN 2-920858-00-2